



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 169186

TO: Jeanine Goldberg  
Location: REM-2D15&2C70  
Art Unit: 1634  
Monday, October 31, 2005  
Case Serial Number: 10/654416

From: Toby Port  
Location: Biotech-Chem Library  
REM-1A59  
Phone: 571-272-2523

[toby.port@uspto.gov](mailto:toby.port@uspto.gov)

### Search Notes

Examiner Goldberg,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Toby Port  
X22523

***This Page Blank (uspto)***



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact*:

Mary Hale, Information Branch Supervisor  
Remsen Bldg. 01 D86  
571-272-2507

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library, Remsen Bldg.



***This Page Blank (uspto)***



STIC-Biotech/ChemLib

169/86

From: Goldberg, Jeanine  
Sent: Thursday, October 20, 2005 2:42 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/654,416- Search

Hello- Please search for a nucleic acid encoding SEQ ID NO: 14.

Please search for SEQ ID NO: 13.

THANK YOU

Jeanine Anne Goldberg  
1634  
571-272-0743  
REM 2D15  
Mailbox: 2C70

CRFS

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

*This Page Blank (uspio)*



241 ATTGGAAGAAATGATCAATCAAGAAACCTTAAGAAAGCAATTAAGAAATTT 300  
301 CAGATTTAAACGATTAAGATTTGCAAGAGAAATGAACAACAATGCTATTTTGGAGGA 360  
301 CAGATTAATACGATTAAGATTTGCAAGAGAAATGAACAACAATGCTATTTTGGAGGA 360  
361 GTATGTTCAAAAGAACTTAAGAGATTTGCAATTAAGATTTGCAAGATTAAGATTTTGA 420  
361 GTATGTTCAAAAGAACTTAAGAGATTTGCAATTAAGATTTGCAAGATTTTGA 420  
421 AAACGTTGTTAAAGTTGCGGAGAAAGTTCTTTTAAAGAGACCTTGAGGATTTTAAATGA 480  
421 AAACGTTGTTAAAGTTGCGGAGAAAGTTCTTTTAAAGAGACCTTGAGGATTTTAAATGA 480  
481 TCAGTCATATGTAAGAAACCTTAAGAGATTTGCTGTCATGCGGAGGAAAGTAT 540  
481 TCAGTCATATGTAAGAAACCTTAAGAGATTTGCTGTCATGCGGAGGAAAGTAT 540  
541 GAGTTAACAAACTTGTGTCTGAACAGAAAGACATGTAAAGATATTTTAAATGA 600  
541 GAGTTAACAAACTTGTGTCTGAACAGAAAGACATGTAAAGATATTTTAAATGA 600  
601 GATTAAGAGTGGGATCTTTTAAACAGATGTTTCAAGAGCATGGAAGTTTAA 660  
601 GATTAAGAGTGGGATCTTTTAAACAGATGTTTCAAGAGCATGGAAGTTTAA 660  
661 GAAACATGCTTGAATTAAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAGAC 720  
661 GAAACATGCTTGAATTAAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAGAC 720  
721 GATATTAATTAATGTAATTAAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAG 780  
721 GATATTAATTAATGTAATTAAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAG 780  
781 CCACAGAGACCGGATTTTGAATCAACTAGCCAGAGGCTTCAATGACAGAGATTAAG 840  
781 CCACAGAGACCGGATTTTGAATCAACTAGCCAGAGGCTTCAATGACAGAGATTAAG 840  
841 CTGGAAGAGTTTATTAAGAGATTAAGAGATTTTCAATTTTCAATTTTGGAGAGATCTA 900  
841 CTGGAAGAGTTTATTAAGAGATTAAGAGATTTTCAATTTTCAATTTTGGAGAGATCTA 900  
901 AGAGATGCGACAGCTTTGTTGCAATTTGATCCAAAGATTTCAATTTTCAATTTTGA 960  
901 AGAGATGCGACAGCTTTGTTGCAATTTGATCCAAAGATTTCAATTTTCAATTTTGA 960  
961 GACAAAGAGAAATGCGAAGAGCCCTTCAAAAGCTGCAAAATCTTCATGAACATGAG 1020  
961 GACAAAGAGAAATGCGAAGAGCCCTTCAAAAGCTGCAAAATCTTCATGAACATGAG 1020  
1021 GCTTTAGAAAGTTTATTAAGAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAG 1080  
1021 GCTTTAGAAAGTTTATTAAGAGATTTGCAAGATTTTCAATTTTCAATTTTGGAGAG 1080  
1081 GAATTTGCAAAATGATTAATTAACAACTTGCAAAATTTTCACTTCAAAAGCTCAATATAT 1140  
1081 GAATTTGCAAAATGATTAATTAACAACTTGCAAAATTTTCACTTCAAAAGCTCAATATAT 1140  
1141 CGTCTTTTGTATTCGAACAAAGAAATTAATGAATTTGTTGATGCGAAGGCTTGCAACA 1200  
1141 CGTCTTTTGTATTCGAACAAAGAAATTAATGAATTTGTTGATGCGAAGGCTTGCAACA 1200  
1201 TTTCTTGAAGCAAGATTTGCGAAGATTTGAGTCTTATTTTCAATTTTGA 1260  
1201 TTTCTTGAAGCAAGATTTGCGAAGATTTGAGTCTTATTTTCAATTTTGA 1260  
1261 TGTCCAGATGAGAAATGATTAATTAAGAGATTTTCAATTTTGA 1320  
1261 TGTCCAGATGAGAAATGATTAATTAAGAGATTTTCAATTTTGA 1320  
1321 GATGACGGGCAAAATTAAGTGTCTGCAAGAAATTAAGAGATTTTCAATTTTGA 1380  
1321 GATGACGGGCAAAATTAAGTGTCTGCAAGAAATTAAGAGATTTTCAATTTTGA 1380

1321 GATGACGGGCAAAATTAAGTGTCTGCAAGAAATTAAGAGATTTTCAATTTTGA 1380  
1381 AAAAGTGGCTTGAAGAGTTTCAACAGAAATTAAGATTAAGATTAAGATTAAGATTAAG 1440  
1381 AAAAGTGGCTTGAAGAGTTTCAACAGAAATTAAGATTAAGATTAAGATTAAGATTAAG 1440  
1441 GAAACCAAGAGATTTCTCAACAGATTAATTTATTTATTTATTTATTTATTTATTTATTT 1500  
1441 GAAACCAAGAGATTTCTCAACAGATTAATTTATTTATTTATTTATTTATTTATTTATTT 1500  
1501 GCAGCCGGGTTGCTTACATGATCTTGAATGAAGAAATTAATTTATTTATTTATTTATTT 1560  
1501 GCAGCCGGGTTGCTTACATGATCTTGAATGAAGAAATTAATTTATTTATTTATTTATTT 1560  
1561 GATCAAAAGCGAGATTTCCGACAGATTAATTAAGAGATTTTGAAGATTTGCGAGAGATTC 1620  
1561 GATCAAAAGCGAGATTTCCGACAGATTAATTAAGAGATTTTGAAGATTTGCGAGAGATTC 1620  
1621 GATTTAGAGAGATTTCAAAAGAAATTAATTAAGATTTTGAAGATTTTGAAGATTTTGA 1680  
1621 GATTTAGAGAGATTTCAAAAGAAATTAATTAAGATTTTGAAGATTTTGAAGATTTTGA 1680  
1681 AATGCTTGGGAGCTTACAGAAATTTTAAAGCAGGTTTATTTATTTATTTATTTATTTAT 1740  
1681 AATGCTTGGGAGCTTACAGAAATTTTAAAGCAGGTTTATTTATTTATTTATTTATTTAT 1740  
1741 TTGAAGACCAAGAGATTTGTAATTAATTAAGAGATTTTGAAGATTTTGAAGATTTTGA 1800  
1741 TTGAAGACCAAGAGATTTGTAATTAATTAAGAGATTTTGAAGATTTTGAAGATTTTGA 1800  
1801 AGAGAGATGACCGTTTCTCTTTTGTATGTTCTTTCCAAAGCGCTACGTTGAGCTGATG 1860  
1801 AGAGAGATGACCGTTTCTCTTTTGTATGTTCTTTCCAAAGCGCTACGTTGAGCTGATG 1860  
1861 GTAAAGACGTAAGAGACAGGTTGTAATTAATTAAGAGATTTTGAAGATTTTGAAGATTT 1920  
1861 GTAAAGACGTAAGAGACAGGTTGTAATTAATTAAGAGATTTTGAAGATTTTGAAGATTT 1920  
1921 ATTGAATTTCTTGAAGAAATTAATTAAGATTAATTAAGATTTTGAAGATTTTGAAGATTT 1980  
1921 ATTGAATTTCTTGAAGAAATTAATTAAGATTAATTAAGATTTTGAAGATTTTGAAGATTT 1980  
1981 TGGCATACGATTTGCAATTAATTTTCACTTAATTTGTCAGAGCTTTTACGAAGAGATTA 2040  
1981 TGGCATACGATTTGCAATTAATTTTCACTTAATTTGTCAGAGCTTTTACGAAGAGATTA 2040  
2041 TGTCAAAAGATCAAGAGAGATGAGCGGCTTATTAAGAGAGAGGCTTGAAGAGATGCT 2100  
2041 TGTCAAAAGATCAAGAGAGATGAGCGGCTTATTAAGAGAGAGGCTTGAAGAGATGCT 2100  
2101 CTCAAGTGAAGCTTCAAGAGAAATTTGAATGATTAATTTGAATGATTAATTTGAATTTGA 2160  
2101 CTCAAGTGAAGCTTCAAGAGAAATTTGAATGATTAATTTGAATGATTAATTTGAATTTGA 2160  
2161 AGATTTTGTACAGTACGAGAGAAACGTAATTAATGCTGCAATCAAGTGTATGCAAACT 2220  
2161 AGATTTTGTACAGTACGAGAGAAACGTAATTAATGCTGCAATCAAGTGTATGCAAACT 2220  
2221 AAACCAAGAGATTAATCTGGAAGAGATGAGATGCTTGAAGAGAGAGCTCTGTAAGAA 2280  
2221 AAACCAAGAGATTAATCTGGAAGAGATGAGATGCTTGAAGAGAGAGCTCTGTAAGAA 2280  
2281 TTAAGTGAAGAGATGGAAGAACAGTGAAGCAATTAAGCAAGATTTGAAGCAACCGGCA 2340  
2281 TTAAGTGAAGAGATGGAAGAACAGTGAAGCAATTAAGCAAGATTTGAAGCAACCGGCA 2340  
2341 GCTGATCTTAAAGAAATTAATTAAGATTAATTAAGATTTTGAAGCAAGATTTGAAGCA 2400  
2341 GCTGATCTTAAAGAAATTAATTAAGATTAATTAAGATTTTGAAGCAAGATTTGAAGCA 2400  
2401 ATGAACAGTCCAGCTTTGTTGTCATCTCATTAAGAGAGAGAGATTAATTAATTAATTA 2460  
2401 ATGAACAGTCCAGCTTTGTTGTCATCTCATTAAGAGAGAGAGATTAATTAATTAATTA 2460

QY 2461 AGTAATAGCAAAAAAGATAGATGCGCTTCAAAAGGACTTCAAGATACCAAAAA 2520  
 DB 2461 AGTAATAGCAAAAAAGATAGATGCGCTTCAAAAGGACTTCAAGATACCAAAAA 2520  
 QY 2521 CATGTGAAAAATCTACGAGAGGGAGTTAAGATGATCCGTACAGAAATTGAAGCTAA 2580  
 DB 2521 CATGTGAAAAATCTACGAGAGGGAGTTAAGATGATCCGTACAGAAATTGAAGCTAA 2580  
 QY 2581 GCATTGATTTGGCAGCAGAAAGTATTTGGAAGATATGATGTTGAAGAAAGTATAT 2640  
 DB 2581 GCATTGATTTGGCAGCAGAAAGTATTTGGAAGATATGATGTTGAAGAAAGTATAT 2640  
 QY 2641 AAATTGGAATCAGATTGCAAGATTAGGAGGATTGCAAAAGACTTGAAGAAAGTATGCAA 2700  
 DB 2641 AAATTGGAATCAGATTGCAAGATTAGGAGGATTGCAAAAGACTTGAAGAAAGTATGCAA 2700  
 QY 2701 AAGATTATATAGAGCTTGTCTGCAATCTGAAGCCTCTGAGAGTGAAGCCGACGAAACAGTG 2760  
 DB 2701 AAGATTATATAGAGCTTGTCTGCAATCTGAAGCCTCTGAGAGTGAAGCCGACGAAACAGTG 2760  
 QY 2761 ACAGAAAGTACAGACGACATCTACAAACAAACAAACCCGTTGCCGATCCGAAAGCAACG 2820  
 DB 2761 ACAGAAAGTACAGACGACATCTACAAACAAACAAACCCGTTGCCGATCCGAAAGCAACG 2820  
 QY 2821 GAATGCAAAATCCTTACAGACAAAGACATGAGGTTACAGACATGACACACACAAAGC 2880  
 DB 2821 GAATGCAAAATCCTTACAGACAAAGACATGAGGTTACAGACATGACACACACAAAGC 2880  
 QY 2881 ACCTCTACTATCAGATCTACCATCATCATCAAAAATTAACATTGACATCAACGAGCGATGC 2940  
 DB 2881 ACCTCTACTATCAGATCTACCATCATCATCAAAAATTAACATTGACATCAACGAGCGATGC 2940  
 QY 2941 AAACCAACCAAGTGTACGACGAGAGAGAGATGATGACGAGACGCTGAAACCGAGTGAG 3000  
 DB 2941 AAACCAACCAAGTGTACGACGAGAGAGAGATGATGACGAGACGCTGAAACCGAGTGAG 3000  
 QY 3001 GGGGTGAGGATGAGGAGGAGTGAATGTGATGAGGAGGAGGATAGTAAAGTATTTTCG 3060  
 DB 3001 GGGGTGAGGATGAGGAGGAGTGAATGTGATGAGGAGGAGGATAGTAAAGTATTTTCG 3060  
 QY 3061 TTTCATGATTTAG 3072  
 DB 3061 TTTCATGATTTAG 3072

RESULT 2  
 AF033211 3071 bp DNA linear PLN 10-SEP-1998  
 LOCUS AF033211  
 DEFINITION glycoprotein (MSG) gene, partial cde.  
 ACCESSION AF033211 GI:3560518  
 VERSION AF033211.1  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pneumocystis jirovecii  
 Pneumocystis jirovecii  
 Bukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.  
 REFERENCE  
 AUTHORS Garbe, T.R., and Striinger, J.R.  
 TITLE Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii  
 JOURNAL Infect. Immun. 62 (6), 3092-3101 (1994)  
 MEDLINE 94314421  
 PUBMED 751806  
 REFERENCE  
 AUTHORS Mei, Q., Turner, R.E., Social, V., Klivington, D., Angus, C.W., and Kovacs, J.A.  
 TITLE Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region  
 JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)  
 MEDLINE 98380374

PUBMED 9712777  
 REFERENCE 3 (bases 1 to 3071)  
 AUTHORS Mei, Q., Turner, R., Social, V., Klivington, D., Angus, C.W., and Kovacs, J.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA  
 FEATURES  
 source  
 location/Qualifiers  
 1..3071  
 /organism="Pneumocystis jirovecii"  
 /mol\_type="genomic DNA"  
 /specific\_host="Homo sapiens"  
 /sub\_species="hominis"  
 /db\_xref="taxon:42068"  
 /clone="HUMSG35"  
 /note="derived from HIV-infected human with P. carinii pneumonia"  
 <1..3071  
 /gene="MSG"  
 /gene="MSG"  
 /note="surface antigen"  
 /codon\_start=3  
 /product="major surface glycoprotein"  
 /protein\_id="AAC34974.1"  
 /db\_xref="GI:3560519"  
 /translation="AAAVKQAAAGTQNSIDBEHYALILKEDGSEBCEKKKLLKVCY  
 ELTEAKNITQVHRKLFCEDEGADPTNCEKLANIEKCTTIGKLEIKKIQIT  
 TDQCKENBQCEFLBEGVCKELKDDNTLNKCYORAKVVEVILRLRLSDNS  
 VICEKIKKEICPMVGRSDELTYNLCLNKECTCKNILLKXPKCGTLTVDYSAAGSRK  
 KENCLELBEOCYFYIGNCGDDIICLELKGCOEONIAVMPPEPDPRPRAETIAE  
 DIGLEFRYKVEEDGVPIKQNLBDNALIALIIDSLLKKDKCECEALCKEY  
 PHEBHALESCKKGLSNDGTCKCELENDINTCKIFSKVTNNRLPRPKNNBY  
 GMEGLPTFLSNEDCAKLESYCFPEKCEPGENACKNIRATCRKGLDAPANKVLQEN  
 MRGLHNSNSWLEKFOELVYKCEKCKEKNKSFSNDELFLCVQAKAARLLTHL  
 RKMTIFLRQOLDQKDPIDPKCKELRKCDLGEISKEITWPCHTLEOCCNRIGTTE  
 ILKQVLDEHKDILKDOESCCKYLKCKNKSRRGDRFSFVCFQNAETELMWKDYK  
 DRCEVFKNIKASYIIEFLNNNTKLTLENNCPSMHTYCNRFSPNCGRLTKENSCKR  
 IKHREBPYRKALIEDLKYELQKLLTDSKCEPALRKYCTTVAGNVNNAISGLCKAN  
 IKONSGKSDERKELECEKLVKEVECEKALPELGGPADLKDQVYTTBELKRAE  
 AMNKSILVSLIKNENSVSKNSKNCKNAVSGLQDPTTKHVILIRGVADVSVEL  
 EAKAFDLAAVEFGYVDLKEKCNKLESDCKIKEDCKDLSEVCKKINKACNKLKLEVK  
 PHEVTSESTTTTTTTTTTVAADPKATECKSLQTTDTWTQTSSTSTSTSTSTSTST  
 LTRSTRCKPTKCTTGGEDDAGDVKPSGLMSGNVWRGVIVAVVISFMT"  
 ORIGIN  
 Query Match 100.0%; Score 3071; DB 8; Length 3071;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TGGCGCGGGCGGTAAACCGGACAGGACACAGATGATGATGAGAAACATG 61  
 DB 1 TGGCGCGGGCGGTAAACCGGACAGGACACAGATGATGATGAGAAACATG 60  
 QY 62 TTTTACCTTAAATCTAAAGGAAGATGAGACTAAGTGAAGCAATGCAAAAAAACTAA 121  
 DB 61 TTTTACCTTAAATCTAAAGGAAGATGAGACTAAGTGAAGCAATGCAAAAAAACTAA 120  
 QY 122 AAAAATATTTGCAAGATTGACTGAACAAAACCTAAATATAGAACAAATACAGAAAAC 181  
 DB 121 AAAAATATTTGCAAGATTGACTGAACAAAACCTAAATATAGAACAAATACAGAAAAC 180  
 QY 182 TTTAAAGTTTTTGGGAAGATGAAAAACAGATACAAATAATGCAAGAACTGAAGCCCATAT 241  
 DB 181 TTTAAAGTTTTTGGGAAGATGAAAAACAGATACAAATAATGCAAGAACTGAAGCCCATAT 240  
 QY 242 TTGAGAAAAAATGACTACATCAAGAAAACTTAAGAAGCAATTAATAAAAAATTC 301  
 DB 241 TTGAGAAAAAATGACTACATCAAGAAAACTTAAGAAGCAATTAATAAAAAATTC 300  
 QY 302 AGATTATACGATTAAGATTGCAAGAGATGAACAACATGCTTATTTTGGAGGAG 361

Dd 301 AGATTATAACGATAGAGATTGCCAAGAGATGAACAAACATGCTATTTTGGAGGAG 360  
Oy 362 TATGTTCAAAAGACTTAAAGATGATTCACATCTTTGAGAAATTAAGTCTATCAAAAGA 421  
Dd 361 TATGTTCAAAAGACTTAAAGATGATTCACATCTTTGAGAAATTAAGTCTATCAAAAGA 420  
Oy 422 AACGTGATTAAGTTCCGAGAGAGATTCTTTAAGAGCATTCGTAGCCATCTTAATGGAT 481  
Dd 421 AACGTGATTAAGTTCCGAGAGAGATTCTTTAAGAGCATTCGTAGCCATCTTAATGGAT 480  
Oy 482 CAGTATATGTAAGAAAAAACTTAAGAGATTGGCCGTGCACTGGGGAGGGAAAGTGATG 541  
Dd 481 CAGTATATGTAAGAAAAAACTTAAGAGATTGGCCGTGCACTGGGGAGGGAAAGTGATG 540  
Oy 542 AGTTAAACAACTTGTGTCTGAACAGAAAGAGACATGTAGAAATATTTAAATTGAAAAAG 601  
Dd 541 AGTTAAACAACTTGTGTCTGAACAGAAAGAGACATGTAGAAATATTTAAATTGAAAAAG 600  
Oy 602 ATAGAGATGCCGTACTCTTTAAACAGATGTTCAAGACACTAGAAAGTTTAAAAAG 661  
Dd 601 ATAGAGATGCCGTACTCTTTAAACAGATGTTCAAGACACTAGAAAGTTTAAAAAG 660  
Oy 662 AAACATGCTTGAATTAAGTCTGCAACAAATGCTATTTTACATGGAAATTCGGAGAGAG 721  
Dd 661 AAACATGCTTGAATTAAGTCTGCAACAAATGCTATTTTACATGGAAATTCGGAGAGAG 720  
Oy 722 ATATATTAATTAATGATTTGAATTGGAGGAAAAATGCCAAGAACAAATATTTGCTATATGC 781  
Dd 721 ATATATTAATTAATGATTTGAATTGGAGGAAAAATGCCAAGAACAAATATTTGCTATATGC 780  
Oy 782 CACCAAGAACCCGATTTTGAATCCACTAGGCCAGAGGCTTCAATAGCAGAGATATAGGCG 841  
Dd 781 CACCAAGAACCCGATTTTGAATCCACTAGGCCAGAGGCTTCAATAGCAGAGATATAGGCG 840  
Oy 842 TGGAGAGATTTAATAAGAGGTAGAGAGATGAGATTTTTATGGAAAGATCATCTTA 901  
Dd 841 TGGAGAGATTTTAATAAGAGGTAGAGAGATGAGATTTTTATGGAAAGATCATCTTA 900  
Oy 902 GAGATGCAACAGCTTGTGGCATTTGATTCACAAATTCATAGTCTTAATAAAAAAGACG 961  
Dd 901 GAGATGCAACAGCTTGTGGCATTTGATTCACAAATTCATAGTCTTAATAAAAAAGACG 960  
Oy 962 ACAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAACTGAGG 1021  
Dd 961 ACAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAACTGAGG 1020  
Oy 1022 CTTTGAAGATTATGTAAGAAAAATGGTTAAGTATGATGGAAGCAAAAAATGTGAG 1081  
Dd 1021 CTTTGAAGATTATGTAAGAAAAATGGTTAAGTATGATGGAAGCAAAAAATGTGAG 1080  
Oy 1082 AATTGCAAAATGATATTAACAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATAATC 1141  
Dd 1081 AATTGCAAAATGATATTAACAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATAATC 1140  
Oy 1142 GTCTTTTGTGATCAACAAAGGAAATTAATGAATTTGTTGATGGAAGGCTTCCCAAT 1201  
Dd 1141 GTCTTTTGTGATCAACAAAGGAAATTAATGAATTTGTTGATGGAAGGCTTCCCAAT 1200  
Oy 1202 TTCTTAGCAAGAGATTGTGGAAATTTGAGATCCATTTGTTCTAATTTTGAAAAAAAT 1261  
Dd 1201 TTCTTAGCAAGAGATTGTGGAAATTTGAGATCCATTTGTTCTAATTTTGAAAAAAAT 1260  
Oy 1262 GTCCAGATGAGAAAAATGCAATTAATAATTAAGAGCAATGTTAACAAGAGAGACTTG 1321  
Dd 1261 GTCCAGATGAGAAAAATGCAATTAATAATTAAGAGCAATGTTAACAAGAGAGACTTG 1320  
Oy 1322 ATGCAACGGGCAATTAAGTCTGCAAGAAATATGCGAGGAATGTTACATGTTCAAAACA 1381  
Dd 1321 ATGCAACGGGCAATTAAGTCTGCAAGAAATATGCGAGGAATGTTACATGTTCAAAACA 1380  
Oy 1382 AAAGCTGGCTTGAAAAAGTTTCAACAAGAAATTAGTAAAGTATGTAAGAAATGAAAAAG 1441  
Dd 1381 AAAGCTGGCTTGAAAAAGTTTCAACAAGAAATTAGTAAAGTATGTAAGAAATGAAAAAG 1440

Oy 1442 AAAACAAAGAGTTTCTCAACGATGAATTAATTAATCTGTGTGTACAGCCAGCAAAAG 1501  
Dd 1441 AAAACAAAGAGTTTCTCAACGATGAATTAATTAATCTGTGTGTACAGCCAGCAAAAG 1500  
Oy 1502 CAGCCGGTGTCTTACACATGATCTTGGAAATGAAAAATATCTTTTTCGACAAACAATCG 1561  
Dd 1501 CAGCCGGTGTCTTACACATGATCTTGGAAATGAAAAATATCTTTTTCGACAAACAATCG 1560  
Oy 1562 ATCAAAAGCAGATTTTCCGACAGATATAATTTGCAAGAAATTTGGGAGAAAGTCCAAAG 1621  
Dd 1561 ATCAAAAGCAGATTTTCCGACAGATATAATTTGCAAGAAATTTGGGAGAAAGTCCAAAG 1620  
Oy 1622 ATTTAGAGAGATTTCAAAAGAAATTAATGAGCCATGTCATACACTGAGAGCAATGCA 1681  
Dd 1621 ATTTAGAGAGATTTCAAAAGAAATTAATGAGCCATGTCATACACTGAGAGCAATGCA 1680  
Oy 1682 ATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTAAAGAGTGAACACAAAGATCTT 1741  
Dd 1681 ATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTAAAGAGTGAACACAAAGATCTT 1740  
Oy 1742 TGAAGAACAAGAAAGTTGTGTAAATTAATCTAAAGAAAGTGTAAATGCTCTAGAA 1801  
Dd 1741 TGAAGAACAAGAAAGTTGTGTAAATTAATCTAAAGAAAGTGTAAATGCTCTAGAA 1800  
Oy 1802 GAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAAACGCTACGTGTGAGCTGATGG 1861  
Dd 1801 GAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAAACGCTACGTGTGAGCTGATGG 1860  
Oy 1862 TAAAGACGTGAAGACAGGCTGAGATTAATCAAAAAAATATTAAGGCTCATATATTA 1921  
Dd 1861 TAAAGACGTGAAGACAGGCTGAGATTAATCAAAAAAATATTAAGGCTCATATATTA 1920  
Oy 1922 TTGAATTTCTGAATAATATCAATTAATAATTAACAACACTGAAAGAAATTTGCCCTT 1981  
Dd 1921 TTGAATTTCTGAATAATATCAATTAATAATTAACAACACTGAAAGAAATTTGCCCTT 1980  
Oy 1982 GGCATACGTATTTGCAATGATTTTACCTAATTTGTCCAGGTCTTAAGAAAGAAATGTT 2041  
Dd 1981 GGCATACGTATTTGCAATGATTTTACCTAATTTGTCCAGGTCTTAAGAAAGAAATGTT 2040  
Oy 2042 GTACAAAATCAAAAGATGTGAGCCGTTCTTAATAAGAAAGCCTTGGAAGATGCTC 2101  
Dd 2041 GTACAAAATCAAAAGATGTGAGCCGTTCTTAATAAGAAAGCCTTGGAAGATGCTC 2100  
Oy 2102 TCAAGTATGAGCTTCAAGAAAAATTTGACTGAATATCTAATGTGAACCTGCAATGAAA 2161  
Dd 2101 TCAAGTATGAGCTTCAAGAAAAATTTGACTGAATATCTAATGTGAACCTGCAATGAAA 2160  
Oy 2162 GATATTTGTACGTAGCCGGAACGTAAATTAATGCGTCAATCAGTGGCTTAAGCAAGCTA 2221  
Dd 2161 GATATTTGTACGTAGCCGGAACGTAAATTAATGCGTCAATCAGTGGCTTAAGCAAGCTA 2220  
Oy 2222 ACACCAAGGATTAATCTGGAAGAGAGTGAAGATGCTAAGAAAGAAATCTGTGAGAAAT 2281  
Dd 2221 ACACCAAGGATTAATCTGGAAGAGAGTGAAGATGCTAAGAAAGAAATCTGTGAGAAAT 2280  
Oy 2282 TAGTGAAGAAAGTGAAGAACAGTGCAAAGCATTAACCAACGAATTTAGAGCAACCGGAG 2341  
Dd 2281 TAGTGAAGAAAGTGAAGAACAGTGCAAAGCATTAACCAACGAATTTAGAGCAACCGGAG 2340  
Oy 2342 CTGATCTTAATAAAAGATTATAGACATATGAGAACTTTAAGAAACGTGCAAGAGAACAA 2401  
Dd 2341 CTGATCTTAATAAAAGATTATAGACATATGAGAACTTTAAGAAACGTGCAAGAGAACAA 2400  
Oy 2402 TGAACAAGTCCAGTCTTGTTTGTCACTCATTTAAGAAAAAAGAAAGTAAATGATCAAAA 2461  
Dd 2401 TGAACAAGTCCAGTCTTGTTTGTCACTCATTTAAGAAAAAAGAAAGTAAATGATCAAAA 2460  
Oy 2462 GTAATAGCAAAACAAAGATTAAGATGCGTTTCAACGAGCTTCAAGATACCAAAAC 2521  
Dd 2461 GTAATAGCAAAACAAAGATTAAGATGCGTTTCAACGAGCTTCAAGATACCAAAAC 2520

Qy	2552	ATGTGAAAATCTACGGAGGGGAGCTTAAAGATGATATCCGTAAACGAATTTAAGACTTAAG	2581
Db	2521	ATGTGAAAATCTACGGAGGGGAGTAAAGATGATATCCGTAAACGAATTTAAGACTTAAG	2580
Qy	2582	CATTGATTTTGGCAGCAGAAAGTATTTTGGAAATATGTAGATTTTGAAGAAAGATGTATA	2641
Db	2581	CATTGATTTTGGCAGCAGAAAGTATTTTGGAAATATGTAGATTTTGAAGAAAGATGTATA	2640
Qy	2642	AATTGGAATCAGATTGCAGAAATTAAAGAGATTGCAAAAGACTTGAAGAAAGTATGCAAA	2701
Db	2641	AATTGGAATCAGATTGCAGAAATTAAAGAGATTGCAAAAGACTTGAAGAAAGTATGCAAA	2700
Qy	2702	AGATTTAATTAAGCCTTGTCCGCAATCTGAAGCCTCTGGAGGTGAACCCGACGAACAGTGA	2761
Db	2701	AGATTTAATTAAGCCTTGTGTGCAATCTGAAGCCTCTGGAGGTGAACCCGACGAACAGTGA	2760
Qy	2762	CAGAAAGTACACAGACAACTACCAACAAACAAACCGCTTGCCGATCCGAAGGCCAAACGG	2821
Db	2761	CAGAAAGTACACAGACAACTACCAACAAACAAACCGCTTGCCGATCCGAAGGCCAAACGG	2820
Qy	2822	AATSCAAATTCCTTACAGACAAACAGACATGGGTTCACAGACATCGACACACAACAGCA	2881
Db	2821	AATSCAAATTCCTTACAGACAAACAGACATGGGTTCACAGACATCGACACACAACAGCA	2880
Qy	2882	CGTCTACATATGACATCTACCAATCCAAATCAATGAACATCAACGAGGGGATGCA	2941
Db	2881	CGTCTACATATGACATCTACCAATCCAAATCAATGAACATCAACGAGGGGATGCA	2940
Qy	2942	AACCAACCAAGTGTACGACAGAGAGAGAAATGTATGCAGAGAGACGTGAAAACCGAGTGAAG	3001
Db	2941	AACCAACCAAGTGTGTACGACAGAGAGAGAAATGTATGCAGAGAGACGTGAAAACCGAGTGAAG	3000
Qy	3002	GGCTGAGAGTAGTGGGTGTGAATGTATGAGGGGGGTGATATGTACCAATGGTTATTTTCGT	3061
Db	3001	GGCTGAGAGTAGTGGGTGTGAATGTATGAGGGGGGTGATATGTATGTATTTTCGT	3060
Qy	3062	TCATGATTTTAA 3072	
Db	3061	TCATGATTTTAA 3071	

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AF033210	3056 bp	DNA	linear	PLN 10-SBP-1998			
		Pneumocystis carinii f. sp. hominis clone HUMS03 major surface glycoprotein (MS) gene, partial cds.	AF033210				
			AF033210.1	GI:3560516			
		Pneumocystis jirovecii					
		Pneumocystis jirovecii					
		Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;					
		Pneumocystidaceae; Pneumocystis.					
		1 (bases 1 to 3056)					
		Gashe, T.R. and Stringer, J.R.					
		Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii					
		Infect. Immun. 62 (8), 3092-3101 (1994)					
		94314421					
		7518806					
		2 (bases 1 to 3056)					
		Wei, Q., Turner, R.E., Social, V., Livingston, D., Angus, C.W. and Kovacs, J.A.					
		Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region					
		Infect. Immun. 66 (9), 4268-4273 (1998)					
		98380374					
		9712777					
		3 (bases 1 to 3056)					
		Wei, Q., Turner, R., Social, V., Livingston, D., Angus, C.W. and Kovacs, J.A.					
		Direct Submission					

JOURNAL	Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA
FEATURES	Location/Qualifiers
Source	1. 3056 /organism="Pneumocystis jirovecii" /mol_type="genomic DNA" /specific_host="Homo sapiens" /sub_species="hominis" /db_xref="taxon:42068" /clone="HumSg33" /notes="derived from HIV-infected human with P. carinii pneumonia"
gene	<1. 3056 /gene="MSG"
CDS	<1. 3056 /gene="MSG" /note="surface antigen" /codon_start=3 /product="major surface glycoprotein" /protein_id="AAC34973.1" /db_xref="GI:3560517" /translation="AAVAKKQVVGASQYDDEVNIALILQEDAMEDTKCKSLKRC EELKRLSIDMEKHVKLKDRCGNGKASKATKCGAOKATGTCNKTOKLPALTN PSDNCKESBRQCLFLGACNHYEDCNKRLNLCYOKKPDGAEVLVLRSLDINT ETHEKLUKEICPVLORESNSLTDICLNLQKTCENIILEKDKCTTLKAVATLALGSK KEICLBLEQCVYVIGNCGDDI1KICIELQKOEONIVYIPGPBPDPRPEVTLAE DISDHEFLKAEEDGVFIKGHMLRDATALTTLVYKDDQTKNNKIGKCKIIEEDCK NIOOHEALKRLCANNSPNAYGKCKEKELEIDIKTCNCKLPTLTKHLVYPNKIVEM RLRPTLVNEDKARLESYCPYVERKACPAKEBECMI1RAACVYRGDGRANKVLOENNR GLIRGNSQSLWKPQRLVYVCELEKENSFPNDEIFVLVCVPAKARLLTHDHQNR VILRQQLDQKDPFDKDKCKELGKCCODLGKSKETWPCHTLPEQCNLGLTEILK OVLDEHKDILKDOESCVKYLKEKCNMSRRGRDRESFVCFQNAITCELVKQVRC EKKRNLKASYIIEPLENNNTKLTTLERNCPSHATCNPSFNCPLGTENSCTKILKK HCEPFRKALIEDALVKELQGLTKDSKCEPALKRYCTVAGVNNMSISGLCRANTDQ NSKSLVSLIKENESNVSKSNSKMKDKNANSNGI0DTTKHYKILIRGKDVSVTELEAK AFPLAEVFRVYDLKERCNKLSIDRCIKDDCKDEVCCKINAKCNLKPLEKPHIE TVLESITTTTTVTYVADPRATECKSLQTDCTVTOSTHSTSTSTSTSTSTSTSTST TRCKPKTKCTGDDAEDVKKPSBGLRVSGNMVMSGVIAVMISPMI"
ORIGIN	
Query Match	75.5%; Score 2320.8; DB 8; Length 3056;
Best Local Similarity	86.0%; Pred. No. 0;
Matches 2646; Conservative	0; Mismatches 404; Indels 27; Gaps 6;
QY	2 TGGCGCGCGGGGCTCAAGCGGCAGG---CACAGGGGACACAGAAATAGCATTTGATGAGGAC 58
Db	1 TGGCGCGCGGGGCTCAAGCGGCAGGTAACAGAGAGATCAGGCGCAATATGATGATGAATGA 60
QY	59 ATGTTTACGCTTTATATCTTAAAGGAAGATGAACTAAGTGAGCAGGAATGCAGAAAAAAAC 118
Db	61 ATATTTTGGCGGTGATTCTACAGAGAGATGTCATGAAAGATACAAATTCGCAAAAAAGTT 120
QY	119 TAAAAAATAATTGCGAAGAAATGACTGAGACAAAACCTAAATATGAAACAAGTACACAGAA 178
Db	121 TAGAAAAATATCTGGAGAGAGTTGAAAAAACATCATCTAGACATGAGAAAAAGTACATAAA 180
QY	179 AACTTAAAGTTTTTGGCAAGATGCAAAACGACATACAAATGCAAAAGCATGAAACCA 238
Db	181 TGCTTAAAGTTTCTTGAGAAATGGGAAACCAAGTAAGCAAAATGCATMAAGTC 240
QY	239 ATATTGAGAAAAATGTACTACATCAAGAGAAAACTTAAAGAACGCAATTAATTAATAAAAA 298
Db	241 TACAAGCCAAAGTTACGGGGAAATGTACAAATTTTAAACAACAAAGCTAGACACAGCT 300
QY	299 TTCAGATTATACGGAATAGGATTGCAAGAAGATGAACAACAATGCTATATTTTGGAG 358
Db	301 TAACAATTCATCGATGATTAATTGCAAAAGAGATGAACGACAAATGCTATTTTGGAGG 360
QY	359 GAGTATGTTCAAAAAGAACTTAAAGATGATTTGCAATACCTTTGAGAAAAATAAATGCTATAAA 418
Db	361 GAGCATG---CCATTAATCTTGTGAAAGAAATGTGTAAACAACCTTAAGGAATCTATGTTTACGAA 417



Oy	413	GGAAACGATAAAGCTTCGGGAAGAAGTCTTTTAAAGCACTTCGATGCACTCTTAATG	478
Db	418	AAAAACGTGACGAGTAGCAGAGAAGTCTTTTGAAGGCACTTCGATGATCTCAATA	477
Oy	479	GATCAGTCATATGTGAAAAAAAACCTTAAGAATTTGGCCTGATGGGAGGAAAGTG	538
Db	478	AAACAGAAACATGAAAAAAAACCTGAAGAAGATTGGCCAGCTTCGACAGGGAAAGTA	537
Oy	539	ATGAGTTAACAACTTGTGTCTGAACCGAAGAAGACATGTAAAGATATTTTAAATTGAA	598
Db	538	ATGATTTAACGACACTGTGTGTGAACCGAAGAAAGCCTGGAGATATTTTAAAGAAA	597
Oy	599	AAGATTAAGAAGCGGTACTCTTTAAACAGATGTTTCAGCACTGAGGAGTTTAA	658
Db	598	AAGATTAAGAAATGCTACTCTTTAAGCAATGTTCACACGACTTGGAAGTTTAA	657
Oy	659	AAGAAACATGCTCTGAAATTAACCGAACATGCTATTTTACATGGAATTCGAGACG	718
Db	658	AAGAAATATGCTTGAATTAACCGAACATGCTATTTTACATGGAATTCGAGACG	717
Oy	719	ACGATATATTAATGTATTGGAATTGGAGGAAAATGCCAAGAACAAATTTGCTTATA	778
Db	718	ACGATATATTAATGAATGATTGAGAGGAAATGCAAGAACAAATTTGTTATA	777
Oy	779	TGCCACACGAGACCCGATTTTGATCCAACTAGGCGAGGCTACAAATACAGAGATATAG	838
Db	778	TACCAACGAGACCCGATTTTGATCCAACTAGCAGAGGCTACACTAGCAAGGACATAG	837
Oy	839	GGCTGGAAGATTTTATTAAGAAAGTAGAGAGATGAGATTTTATTTGGAAGAATCATC	898
Db	838	ACCTGATGAGCTTTATATAAAGGACAGAAAGAGATGGTGTTTTATGGAACATCATT	897
Oy	899	TAAAGATGCCACAGCTTTGTGGCATGTG--ATCCAGATTCTAGTCTTAAAAAA	955
Db	898	TAAAGATGACACACACTTATTGACGTGTGTTAAGAAAGATGATCAGAAAAATA	957
Oy	956	AAGACGACAAAGAAGAAATGCGAAGAGCCCTCAAAAAAGCTGCAAAATCCTATGAC	1015
Db	958	ATAATATGCGAGAAAAATGCAATAGATCTTCGAGATTAATGCAAAAACCTCACAACG	1017
Oy	1016	ATGAGGCTTTAGAAAATTATGTAGAAGAAAATGTTTAAGTAAATGATGGAACGAAAAAT	1075
Db	1018	ATGAGGCTCTAAAAAATTTATGTATAATTAATTAATGCTCTTAAGCATATGGAAGAAAAAT	1077
Oy	1076	GTGAAGAAATTGCAAAATGATATTTAACAAACTTGCAAAATTTTCCTCAAAAGTCACTA	1135
Db	1078	GCAAGAAATTGMAAMAMATATTAAAAAACATGCAAAACCTCAACCAACGATCTTA	1137
Oy	1136	ATATATCGTCTTTTGTATCCAAACAAAGAAATTAATGAATTTGTTGAGAGGGGTGC	1195
Db	1138	AAATCATCTTTATATATC-----AAATGATTAATTTGTTGAGAGGAAACCTGC	1188
Oy	1196	CAACATTTCTTAGCAACGAGATTTGCGAAATTTGGAGTCCATATGTTCTATTTTGA	1255
Db	1189	CAACATTTCTTACTAATGAAGCTGTGCAAGATTTGAATCTTATGTTTTTACTACGAA	1248
Oy	1256	AAAAATGTCCAGATGGAAGAAATGCAATGTAAAAATTAAGACCAACATTTACAAAGAG	1315
Db	1249	AAGCTGTTCAAATATCCAAAGAAAGTATAGAAATTTAAGGCGACCGTGTATTAAGAG	1308
Oy	1316	GACTTGATGACGCGCAATTAAGTGTGCAAGAAATATGCGAGGAATGTTACATGGTT	1375
Db	1309	GGCTTGATGACGCGCAATTAAGTGTGCAAGAAATATGCGTGGATTAATACGTGGTT	1368
Oy	1376	CAACAAAGCTGCTTGAAGAAATTTCACAAAGTAATGTAAGTAATGTGAGAACTGA	1435
Db	1369	CAAACTCAAGTTGGCTTAAGAGTTTCAACAAAGATTGTAAGTAAGTATGTAAGAGCT--	1426
Oy	1436	AAAAAGAAACAAAGGAATTTCAACAGATGTAATTTATTTCTGTGTGACAGCCAG	1495
Db	1427	AAAAAGAAATTAAGGAATTTCCCAACAGATGAATATTTGTTCTGTGTGACAGCCAG	1485

QY	1496	AAAAGACGCCGGTTCCTTAACCTGATCTCTGAAATGAAAACATCTCTTTTTCACGACAC	1555
Db	1486	CAAAAGCTGCACGATTACTTAACAACATCATCAAAATGAGGGTTATCTTTTTCACGACAC	1545
QY	1556	AACCTGATCTAAAAGCCGAGATTTTCCCGACATATAAAATTTGCCAAGAAATTTGGGCGAAGAT	1615
Db	1546	AAATTGGATCAAAAGAGAGATTTTCCGACAGATTAAGCTCCGAAAGAAATTTAGGGAAAAAT	1605
QY	1616	GCCAAAGATTTAGAGAGAGATTCAAAAGAAATTAATGAGCGATGTATATCACTGAGCAGC	1675
Db	1606	GCCAAAGATTTAGAAAAGATTCAAAAGAAATTAATGAGCGATGTATATCACTGAGCAGC	1665
QY	1676	AATGCATTCGTTGGGAGCTTACAGAAATTTTAAAGCGTTTTATTTGAGTGAACACAAG	1735
Db	1666	AATCCAAATCGCTGGGAGCTTACAGAAATTTTAAAGCGTTTTATTTGAGTGAACACAAG	1725
QY	1736	ATACCTTTGAAAGACCAAGAAAGTGTTGTAAATATCTATPAAAGAAAGTGTATATAATGT	1795
Db	1726	ATACCTTTGAAAGACCAAGAAAGTGTTGTAAATATCTATPAAAGAAAGTGTATATAATGT	1785
QY	1796	CTAGAAGAGAGATGACCGTTTCTCTTTTGTATGTCTTCCAAAACGCTACGTTGAGC	1855
Db	1786	CTAGAAGAGAGATGACCGTTTCTCTTTTGTATGTCTTCCAAAACGCTACGTTGAGC	1845
QY	1856	TGATGCTPAAAAGACTGAAGAACAGGTGTGAAGTATCAAAAAAAATATAAAGCTTCAT	1915
Db	1846	TGATGCTPAAAAGAGCTGAAGAACAGGTGTGAAGTATCAAAAAAAATATAAAGCTTCAT	1905
QY	1916	ATATTATTTGAATTTCTTGAAAAATATATCAATATAAATAACACACCTGGAAGAAATTTGTC	1975
Db	1906	ATATTATTTGAATTTCTTGAAAAATATATCAATATAAATAACACACCTGGAAGAAATTTGTC	1965
QY	1976	CCTCTTGGCATACGTATTCATATGATTTTCACTTAATTGTCCAGGTCTTACGAAAGAGA	2035
Db	1966	CCTCTTGGCATACGTATTTGCAATGATTTTCACTTAATTGTCCAGGTCTTACGAAAGAGA	2025
QY	2036	ATAGTTGTACAAAATCAAGAGCATGTGTGACCGGTCTATPAAAGAAAGCCCTTGGAG	2095
Db	2026	ATAGTTGTACAAAATCAAGAGCATGTGTGACCGGTCTATPAAAGAAAGCCCTTGGAG	2085
QY	2096	ATGCTCTCAAAAGTAGAGCTTCAAGAAAATTTGATCTGATTAATCTPAAATGTGAACCTGCAT	2155
Db	2086	ATGCTCTCAAAAGTAGAGCTTCAAGAAAATTTGATCTGATTAATCTPAAATGTGAACCTGCAT	2145
QY	2156	TGAAAAGATATTGTACACGTAGCGGAAACGTAATATATGCGTCATACGTGGCTTATAGCA	2215
Db	2146	TGAAAAGATATTGTACACGTAGCGGAAACGTAATATATGCGTCATACGTGGCTTATAGCA	2205
QY	2216	AAGCTAACACCAAGATTAATCTTGAAAAGGTGATGAGATGCTAGAAAGGAATCTGTG	2275
Db	2206	AAGCTAACACCAAGATTAATCTTGAAAAGGTGATGAGATGCTAGAAAGGAATCTGTG	2265
QY	2276	AGAAATTAAGTAAAGAGTGGAAAGAACAGTGCAAAGCATTAACAAAGAAATTTAGACAAC	2335
Db	2266	AGAAATTAAGTAAAGAGTGGAAAGAACAGTGCAAAGCATTAACAAAGAAATTTAGACAAC	2325
QY	2336	CGGCACTGATCTPAAAAAAGATATATTAAGACATATGAGAACTTAAAGAAAGTCCAGAGG	2395
Db	2326	CGGCACTGATTTAAAAAAGATATATTAAGACATATGAGAACTTAAAGAAAGTCCAGAGG	2385
QY	2396	AAGCAATGAACAAGTCCAGTCTTGTCTTGTCACTCATTAAGAAAAAGAAAGTAATGTAT	2455
Db	2386	AAGCAATGAACAAGTCCAGTCTTGTCTTGTCACTCATTAAGAAAAAGAAAGTAATGTAT	2445
QY	2456	CAAAAAGTAATAGCAAAAAACAAGATTAAGATTCGCCGTTTCAAACGGACTTCAAGATACCA	2515
Db	2446	CAAAAAGTAATAGCAAAAAACAAGATTAAGATTCGCCGTTTCAAACGGACTTCAAGATACCA	2505
QY	2516	CAAAAACATGTAAATAATCTACGAGGGGAGTTAAGATGTATCCGCTAACAATAATTAGAG	2575
Db	2506	CAAAAACATGTAAATAATCTACGAGGGGAGTTAAGATGTATCCGCTAACAATAATTAGAG	2565
QY	2576	CTAAAGCATTTGATTTGGCAGCAGAGATATTGGAAAGATATGTAGATTTGAGAGAAAGAT	2635



|||||  
Db CTAAGCATTTGATTTGGCAGCAGAGATATTTGAAGATTTGAAAGTAAAGAT 2625  
Qy 2636 GTAAATAATTGATTCAGATTTGCAATTAAGAGATTGCAAGACTTAGAAGAT 2635  
Db 2626 GTATTAATTGATTCAGATTTGCAATTAAGAGATTGCAAGACTTAGAAGAT 2685  
Qy 2696 GCAAAAAGATTAAAGGCTTGTGCAATCTGAAGCTTGAAGGAGGAGCGGCA 2755  
Db 2686 GCAAAAAGATTAAAGGCTTGTGCAATCTGAAGCTTGAAGGAGGAGCGGCA 2745  
Qy 2756 CAGTGACAGAAAGTACACGACAACTACACACACACACACACACACACAC 2815  
Db 2746 CAGTGACAGAAAGTACACGACAACTACACACACACACACACACACACAC 2805  
Qy 2816 CAAACGAAATGCAATCTTACACACACACACACACACACACACACACACAC 2875  
Db 2806 CAAACGAAATGCAATCTTACACACACACACACACACACACACACACACAC 2865  
Qy 2876 CAAACGAAATGCAATCTTACACACACACACACACACACACACACACACAC 2935  
Db 2866 CAAACGAAATGCAATCTTACACACACACACACACACACACACACACACAC 2925  
Qy 2936 GATGCAACACCAACCAAGTGTACGACAGAGAGAGAGATGATGACAGACG 2995  
Db 2926 GATGCAACACCAACCAAGTGTACGACAGAGAGAGAGATGATGACAGACG 2979  
Qy 2996 GTAGAGGCGCTGAGATGATGCGGTGGAATGTGATGAGGCGGTGATGTA 3055  
Db 2980 GTAGAGGCGCTGAGATGATGCGGTGGAATGTGATGAGGCGGTGATGTA 3039  
Qy 3056 TTTGCTCATGATTTAG 3072  
Db 3040 TTTGCTCATGATTTAG 3056

RESULT 4  
AR438439  
LOCUS AR438439 3054 bp DNA linear PAT 20-FEB-2004  
DEFINITION Sequence 11 from patent US 6664053.  
ACCESSION AR438439  
VERSION AR438439.1 GI:42663294  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3054)  
AUTHORS Kovacs,J.A., Huang,S., Maaur,H., Fiescher,S.H., Gyll,V.J. and Mei,Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG)  
JOURNAL gene of human Pneumocystis carinii  
FEATURES  
SOURCE Location/Qualifiers  
1..3054  
/organism="unknown"  
/mol\_type="genomic DNA"

Query Match 75.5%; Score 2318.8; DB 6; Length 3054;  
Best Local Similarity 86.0%; Pred. No. 0;  
Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;

Qy 181 CTTAAAGTTTTTGCAGAGATGAAAAAGCAGATACAAATATGCAAGATCGAAAGCCAA 240  
Db 181 CTTAAAGATTTCTGTGAAATGGAAGCAAGTAAAGCAATATCAAAATGTCAGAGCTCA 240  
Qy 241 ATTGAGAAAAATGTAATCAATCAAGAAAACTTAAAGAGCAATTAAGAAAAATTT 300  
Db 241 CAAGCCAAAGTTACGGGAAATGTACAAATTTTAAACAAAAAGCTAGACCGCTTA 300  
Qy 301 CAGTTTAAACGATTAAGATTTGCAAGAGATGAACAAATGCTATTTTGAAGGA 360  
Db 301 ACAATTCATCAGATTAATTTGCAAGAGATGAACAAATGCTATTTTGAAGGA 360  
Qy 361 GTATGTTCAAAAGAACTTAAAGATGATGCAATATTTGAAATTAAGTCTATCAAA 420  
Db 361 GCATG---CCATATCTTTGAAAGATTTTACAAATTAAGAAATCTATGTCAGAAA 417  
Qy 421 AAACGTATTAAGTTGCGGAAAGATTTTAAAGACCTTGTAGCATTTAATGA 480  
Db 418 AAACGTATTAAGTTGCGGAAAGATTTTAAAGACCTTGTAGCATTTAATGA 477  
Qy 481 TCAATCATATGTAAGAAAAAACTTAAAGATTTGCGGAGGAGGAGATGAT 540  
Db 478 ACAAGAAACATGTAAGAAAAAACTTAAAGATTTGCGGAGGAGGAGATGAT 537  
Qy 541 GAGTTTAAACAACTTGTGCTGAACAGAAAGACATGTAAAGATATTTAATTTGA 600  
Db 538 GAATTTACGACCTTGTGCTTGAACAGAAAGACATGTAAAGATATTTAATTTGA 597  
Qy 601 GATTAAGAGTGGGATCTTTTAAACAGATGTTTACAGACATGAGAAATTTTAA 660  
Db 598 GATTAAGAGTGGGATCTTTTAAACAGATGTTTACAGACATGAGAAATTTTAA 657  
Qy 661 GAAACATGCTGAAATTAATCTGAAACATGCAATTTTAACTTTGAAATTTGCGAG 720  
Db 658 GAAATATGCTGAAATTAATCTGAAACATGCAATTTTAACTTTGAAATTTGCGAG 717  
Qy 721 GATTAATTAATGATTAATGATTTGAGAGAAATGCAAGAACTTATGATG 780  
Db 718 GATTAATTAATGATTAATGATTTGAGAGAAATGCAAGAACTTATGATG 777  
Qy 781 CCACACAGACCCGATTTTGAATCCATGAGCCAGAGCTTCAATGACAGAGATATG 840  
Db 778 CCACACAGACCCGATTTTGAATCCATGAGCCAGAGCTTCAATGACAGAGATATG 837  
Qy 841 CTGGAAGATTTTAAAGAGTTAGAGAGATGAGATTTTAAAGAAATCATCTA 900  
Db 838 CTGGAAGATTTTAAAGAGTTAGAGAGATGAGATTTTAAAGAAATCATCTA 897  
Qy 901 AGAGATCGACAGCTTTGTTGCAATTTG---ATCCAGATTTCTAGCTTTAAAA 957  
Db 898 AGAGATCGACAGCTTTGTTGCAATTTG---ATCCAGATTTCTAGCTTTAAAA 957  
Qy 958 GACGACAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGA 1017  
Db 958 AATATCGAGAAAAATGCAATTAAGATTTCTGAGATTAAGCAAAAAATCTCATGA 1017  
Qy 1018 GAGGCTTTAAGAAATTTAATGAAAGAAATGTTAATGATGATGAAACGAAAAATGT 1077  
Db 1018 GAACTCTTAAAAAAATTTAATGAAAGAAATGTTAATGATGATGAAAGAAATGT 1077  
Qy 1078 GAAAGATTTGCAAAATGATTTAATCAAAATTTTCACTTCAAAAGTCACTAAT 1137  
Db 1078 AAAGATTTAAGAAAGATTTAATCAAAATTTTCACTTCAAAAGTCACTAAT 1137  
Qy 1138 AATGCTTTTGTATGTCACAAAAAGAAATTAATGAAATTTGTTGAGGAGGATGCCA 1197  
Db 1138 AACATCTTTATGATTC-----AATGATTAATTTGTTGAGGAGGATGCCA 1188  
Qy 1198 ACATTTCTTACGACGAAAGATTTGCGAATTTGAGGCTATTTGTTTGAAGAA 1257  
Db 1189 ACATTTCTTACGACGAAAGATTTGCGAATTTGAGGCTATTTGTTTGAAGAA 1248  
Qy 1258 AATGTCAGATGAGAAAAATGCAATTAATTAAGCAACATGTTTCAAAAGAGA 1317

Db	1249	GCTTGTCCAAATGCCAAAGAGAGTGTAATTGAAAGGCGAGCGGTGTTATTAAGAAGGG	1308
Oy	1318	CTTGATGCACGGGCAATPAAGTGCTGCAGAAATATGCGAGAGATGTTACATGGTTCA	1377
Db	1309	CTTGATGACGGGCAAAATPAAGTGCTGCAGAAATATGCGGGGTTATTACGTGTTCA	1368
Oy	1378	AAACAACCTGGCTTGAAGTTCAACAAGATTAGTAAAGTATGTGADAAACTGAA	1437
Db	1369	AATCAAGTTGGCTTAAAGAGTTTAAACAAGATTAGTAAAGTATGTAAAGAGCT--A	1425
Oy	1438	AAAGAAAAAACAAGAAAGTTTCTCAAAACATGATTAATTTATCTGTGTGTCACGCCACA	1497
Db	1426	AAAGAAATTAAGAAGAGTTTCCCAACGATGAAATTTGTTGTGTGTGTCACGCCACA	1485
Oy	1498	AAAGCAGCCCGGTGCTTACACATGATCTTGCATGAAAACTATCTTTTACGACAA	1557
Db	1486	AAAGCTGACGATTAACACACGATCATCAATGAGGGTTATCTTTTACGACAA	1545
Oy	1558	CTGATCAAAAACGAGATTTCCCGACAGTAAATTTGCAAGGAATTTGGGAGAAAGTGC	1617
Db	1546	TTGGATCAAAAAGAGATTTTCCGACAGTAAAGCTCAAGGAATTTGGGAAAAAAGC	1605
Oy	1618	CAAGATTTAGGAGAGATTTCAAAAGAAATTTACATGGCCATGTCATACCTGAGACGAA	1677
Db	1606	CAAGATTTAGGAGAGATTTCAAAAGAAATTTACATGGCCATGTCATACCTGAGACGAA	1665
Oy	1678	TGCAATCGCTTGGGGCTACAGAAATTTTAAAGCAGGTTTATTTGGATGAACACAAAGAT	1737
Db	1666	TGCAATCGCTTGGGGCTACAGAAATTTTAAAGCAGGTTTATTTGGATGAACACAAAGAT	1725
Oy	1738	ACTTTGAAGAAGCAAGAAAGTTGTGTAAATATCTTAAAGAAAGTGTAAATPAATGGTCT	1797
Db	1726	ACTTTGAAGAAGCAAGAAAGTTGTGTAAATATCTTAAAGAAAGTGTAAATPAATGGTCT	1785
Oy	1798	AGAAGAGAGATGAGCGTTTCTCTTTGTATGTGCTTCCAAAAGCGTACGCTGAGCTG	1857
Db	1786	AGAGAGAGATGAGCGTTTCTCTTTGTATGTGCTTCCAAAAGCGTACGCTGAGCTG	1845
Oy	1858	ATGTGTAAGAAGCGTGAAGACAGAGTGTGAAGTATCAAAAAAATATTAAGCTTCATAT	1917
Db	1846	ATGTGTAAGAAGCGTGAAGACAGAGTGTGAAGTATCAAAAAAATATTAAGCTTCATAT	1905
Oy	1918	ATTATTTGAATTTCTTGAATAATATCAAAATPAATPAACAACCTGGAAGAAATGTGCC	1977
Db	1906	ATTATTTGAATTTCTTGAATAATATCAAAATPAATPAACAACCTGGAAGAAATGTGCC	1965
Oy	1978	TCTTGGCAATGATTTGCAATGATTTTCAACCTATTTGCCAGGCTTACGAAAGAGAT	2037
Db	1966	TCTTGGCAATGATTTGCAATGATTTTCAACCTATTTGCCAGGCTTACGAAAGAGAT	2025
Oy	2038	AGTTGTACAAAATCAAGAGAGCATGTGAGCGTCTTATTAAGAAGAGGCTTGGAGAT	2097
Db	2026	AGTTGTACAAAATCAAGAGAGCATGTGAGCGCTTCTATTAAGAAGAGGCTTGGAGAT	2085
Oy	2098	GCTCTCAAGTAGAGCTTCAAGGAAATTTGACTGATTAATCTTAATGTGAACCTGCATTTG	2157
Db	2086	GCTCTCAAGTAGAGCTTCAAGGAAATTTGACTGATTAATCTTAATGTGAACCTGCATTTG	2145
Oy	2158	AAAAGATATTGTACAGTAGCGGGGAAACGTAAATATTTGGTCAATCAGTGGCTTATGCAA	2217
Db	2146	AAAAGATATTGTACAGTAGCGGGGAAACGTAAATATTTGGTCAATCAGTGGCTTATGCAA	2205
Oy	2218	GCTAACACCAAGATTAACCTGTGAAGAAGTATGAGANTGCTAGAAGAAAGAACTCTGTGAG	2277
Db	2206	GCTAACACCAAGATTAACCTGTGAAGAAGTATGAGANTGCTAGAAGAAAGAACTCTGTGAG	2265
Oy	2278	AAATTTAGTGAAGAAAGTGGAAGAACGTGTCAAGCATTAACCAACGAATTTAGAACACCG	2337
Db	2266	AAATTTAGTGAAGAAAGTGGAAGAACGTGTCAAGCATTAACCAACGAATTTAGAACACCG	2325
Oy	2338	GCAAGTGATCTTAAAAAAGATTTATTAAGCATTTGAGGAACTTAAGAAACGTGCAGAGAA	2397

Db	2326	GCACGTGATTTTAAAAAAGATTATPAGACATATGAGAACTTAAGAAACGTGCAGAGGAA	2385
Qy	2398	GCAATGAAACCAAGTCCAGCTCTTTGTCTCACTCTTAAAGAAAAAGAAAGTAATGTATCA	2457
Db	2386	GCAATGAAACCAAGTCCAGCTCTTTGTCTCACTCTTAAAGAAAAAGAAAGTAATGTATCA	2445
Qy	2458	AAAAGTAATAGCAAAAAAAGAGATPAGAAATGCGCTTCAAAACGAGCTTCAAGATACACA	2517
Db	2446	AAAAGTAATAGCAAAAAAAGAGATPAGAAATGCGCTTCAAAACGAGCTTCAAGATACACA	2505
Qy	2518	AAACATGTGAAAATACTACGAGAGGGAGATTAAAGATGTATCCGTAAAGAAATTGAAGCT	2577
Db	2506	AAACATGTGAAAATACTACGAGAGGGAGATTAAAGATGTATCCGTAAAGAAATTGAAGCT	2565
Qy	2578	AAAGCATTTGATTTGGCAGCAGAAAGTATTGGAAAGATATGTATGATTTGAAGAAAGATGT	2637
Db	2566	AAAGCATTTGATTTGGCAGCAGAAAGTATTGGAAAGATATGTATGATTTGAAGAAAGATGT	2625
Qy	2638	AATTAATTTGGAATCAGATTCAGAAATTAAAGAGGATTTGCAAAAGACTTAAGAAAGTATGC	2697
Db	2626	AATTAATTTGGAATCAGATTCAGAAATTAAAGAGGATTTGCAAAAGACTTAAGAAAGTATGC	2685
Qy	2698	AAAAAGATTAATTAAGGCTTGTCCGAATCTGAAGCCTCTGAGAGTGGAAGCCGCAAGAAACA	2757
Db	2686	AAAAAGATTAATTAAGGCTTGTCCGAATCTGAAGCCTCTGAGAGTGGAAGCCGCAAGAAACA	2745
Qy	2758	GTGACAGAAAGTACACACGACACCTTACACACACACACACACACACACACACACACACAC	2817
Db	2746	GTGACAGAAAGTACACACGACACCTTACACACACACACACACACACACACACACACACAC	2805
Qy	2818	ACGGAATGCAAAATCCTTACACAGACAAACAGACACATGGGTTACACAGACATGACACACACA	2877
Db	2806	ACGGAATGCAAAATCCTTACACAGACAAACAGACACATGGGTTACACAGACATGACACACACA	2865
Qy	2878	AGCAGCTCTACTATCAGATCTTACCATCATCATCAAAAAATTAACATTGAGATCAACAGGCGCA	2937
Db	2866	AGCAGCTCTACTATCAGATCTTACCATCATCATCAAAAAATTAACATTGAGATCAACAGGCGCA	2925
Qy	2938	TGCAAAACCAACCAAGTGTACGACACGAGAGAGAGATATGACGAGAGACGTGAAACCGAGT	2997
Db	2926	TGCAAAACCAACCAAGTGTACGACACGAGAGAGAGATATGACGAGAGACGTGAAACCGAGT	2919
Qy	2998	GAGGGCGTGAAGATGAGATGGGTGGAATCTGATGAGGGGGGTGATATGTAAGCAATGGTATT	3057
Db	2980	GAAAGCTTTGAGAGGAGACCGGGTGAATGTATGAGGGGGGTGATATGTAAGCAATGGTATT	3039
Qy	3058	TCGTTTCATGATTTAG 3072	
Db	3040	TCGTTTCATGATTTAG 3054	
RESULT 5			
LOCUS	AR438437	3084 bp	DNA
DEFINITION	Sequence 7 from patent US 6664053.	Linear	PAT 20-FEB-2004
VERSION	AR438437		
KEYWORDS	AR438437.1 GI:42663292		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 3084)		
TITLE	Kovacec,J.A., Huang,S., Maseur,H., Fischer,S.H., Gill,V.J. and Mei,Q.		
FEATURES	Identification of a region of the major surface glycoprotein (MSG)		
FEATURES	Patent: US 6664053-A 7-16-DEC-2003;		
ORIGIN	Location/Qualifiers		
	1..3084		
	/organism="unknown"		
	/mol_type="genomic DNA"		
Query Match	73.7%	Score 2262.6;	DB 6; Length 3084;

Best Local Similarity 84.7%; Pred. No. 0;  
Matches 2640; Conservative 0; Mismatches .399; Indels 78; Gaps 6;

QY 1 ATGCGCGCGCGCTCAACGGCGAGCGAGCACAGAAATACCATTTGATGAGGACAT 60  
Db 1 ATGCGCGCGCGCTCAACGGCGAGCGAGCACAGAAATACCATTTGATGAGGACAT 60  
QY 61 GTTTTACCTTTAATTTCTAAAGAGATGAGCTAAGTACGAGCAATGCAAAAAAATCTA 120  
Db 61 GTTTTACCTTTAATTTCTAAAGAGATGAGCTAAGTACGAGCAATGCAAAAAAATCTA 120  
QY 121 AAAAAATATTTGCCAGATTTGACTGAAGCAAACTAAATATAGAACAGTACACAGAAA 180  
Db 121 GAAGAAATATTTGCCAAACATTTAACTAATGAGGATTAATCCAGAAAAAGTTTCAAGAAAA 180  
QY 181 CTTAAAGCTTTTTCGAGAGATGAGAAAAGCATACAAAATGCAAAAGAAATGGAAGCCAT 240  
Db 181 CTTAAAGCTTTTTCGAGAGATGAGAAAAGCATACAAAATGCAAAAGAAATGGAAGCCAT 240  
QY 181 TTTAAAGATTTCTGTGTATACGGGAAAAGAAATGTCAAAGATCTTAAAAAAACAAA 240  
Db 241 ATTGAGAAAAAATGTACTACATCAAAAGSAAAATTTAAAGAAATTTAAAAAAAATTT 300  
Db 241 GTCAATCAAAAATGATCAATTAATTTCAAGSAAAATTTCAAAACAGCTGCTAGAAAAAATTT 300  
QY 301 CAGATTATTAACGATTAAGATTTGCAAAAGAGATGAACAACATGCTATTTTGGAGGGA 360  
Db 301 TCAGAAATTAACAGATGAGGATTTGCAAAAAGATGAACAACATGCTATTTTGGAGGGA 360  
QY 361 GTATGTTCAAAAAGAACTTTAAAGATTTGCAATCTTTGAGAAATTAAGTGTATCAAAAG 420  
Db 361 GCAATGTTCAACAGAACTTTAAAGATGACTCAATTAATTAAGAAATTAATCTGTATCAAAA 420  
QY 421 AAACGTATTAAGTTGGGAGAAAGTTCTTTAAGAGCACTTGGTAGCATTTAATGGA 480  
Db 421 GAACGGAACATGTGGCAGAAAGTTCTTTGAGGCGCTTGGTGTGATCTCAATGAA 480  
QY 481 TCACTCATATGTGAAAAAAAATTTAAAGAGATTTGCGCTGTCAATGGGAGGAAAGTAT 540  
Db 481 ACAAAACATGTGAAAAAAAATTTAAAGAGATTTGCCGAAATTAAGAAAGAAAGCGAT 540  
QY 541 GAGTTAAACAACTTGTGTCTGACACGAAAGACATGTAAAGATTTTAAATGAAAAA 600  
Db 541 GAATTAACGAGCTTGTCTTATCAAAAACAAACATGCGTAACTTGTAAACAAAAAGA 600  
QY 601 GATAAGAAAGTGGCTACTCTTAAACAGATGTTTACAGAGCACTAGGAAGTTTAAA--- 657  
Db 601 AAAAGTAAATGTGATCTCTTGAAGAAAGATTTGAAGAGCACTTAAAGAAATGAATTTG 660  
QY 658 AAAGAAACATGTCTTGAATTTACTGAAACAATGCTAATTTTACATTTGAAATTTGCGAG-- 715  
Db 661 CGAGAAAAATGTCTACTAATTTACTTGAAGCAATGTTACTTTCACAGAGGAACTGTGAAGGA 720  
QY 716 -----ACGACATATTAATTAATGT 735  
Db 721 GACAAATCAAAAGTCAATTAACCTAATTAATTAAGACTGCAAAAGATATGTATCCAGAGTGT 780  
QY 736 ATTGAATTTGGAGGAAAAATGCGCAAGAACAAATTTTGTATATATGCGCACGAGCCGAT 795  
Db 781 GATGAATTTAGCAAAAAGTGTGAAAAAGAAATATGTTTATATGATCCAGATCCGAT 840  
QY 796 TTTGATCAACTAGAGCGAGAGCTACTACAATAGCAGAGATATAGGCTGGAAGAGTTTAT 855  
Db 841 TTGATCAACTAAGCGAGAGCTACTACAATAGCAGAGATATAGGCTGGAAGAGCTTTAT 900  
QY 856 AAGAAAGTAGAGAGATGAGTTTATTTAGAAAAGATCATCTTAAGAGATGCAACGCT 915  
Db 901 AAGAGGCGAGAGAGATGAGTTTATTTGTTGGAAGACAAACATGTAAAGATGCAACGCT 960  
QY 916 TTGTTGCAATGTTGATCAAGATTTCTAGTCTTAAAAAAAAGCAAAAGAGAAATGC 975  
Db 961 TTGTTGCACTACT-----TCTTAAGAAACCTTTAAAGAAAGAGATGT 1005  
QY 976 GAAGAGCCCTTCAAAAAGCTGCAAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTA 1035

Db 1006 ATAAAAACCCTTAAAAAAAATCAGCAAAACCTCATGAACATGAGGCTTTAGAAAATCTA 1065  
QY 1036 TGTAAAGAAAAATGTTTAAATAGATGAACGAAAAATGTAAGAAATTTGCAAAATGAT 1095  
Db 1066 TGTAAAGAAAAATTAACCAAGTATGATGAACGAAAAATGTAAGAAATTTGCAAAATGAT 1125  
QY 1096 ATTAAACAAAATTTGCAAAAATTTTCACTTCAAAAATGCACTAATAATGCTTTTGTATCCA 1155  
Db 1126 GTTAACAAAATTTGTAACAGTCTTACATCAACAATTTCTTAAAAACGCTTTTACATTTCA 1185  
QY 1156 ACAAAAGAAATATAGAAATTTGTTGATGGAAGGTTGCCAATTTCTTACAGAACGA 1215  
Db 1186 CC-----TATGAATTTGCGGAATGGGAAATTTACCAATTTCTTATGATGA 1236  
QY 1216 GATTGTGCAAAATTTGAGTCTATTTGTTCTATTTTGAAAAAAATGTCAGATGAGAA 1275  
Db 1237 GATTGTGCAAAAATTAAGATTTTATTTGTTTATTAAGAAATTTGTCAGATGTCAA 1296  
QY 1276 AATGATGTAAAAATATAGAGCAACATGTTACAAAAGAGACTTGTATGACGCGCAAT 1335  
Db 1297 GAAGCTTGTATGATGATGAGGCGCGGTTTACAGAGAGAGGCTGTATGACGCGCAAC 1356  
QY 1336 AAAGTCTGCAAGAAATATGCGAGAAATGTACATGTTTCAAAACAAACGCTTGA 1395  
Db 1357 AGTGTGTGCAAAAATATGCGGTGTTATTAAGTGTTCAAATCAAAATTTGGCTTAAG 1416  
QY 1396 AAGTTTCAACAAGATTAATGTAAGTATGTAAGTAACTGAAAAAGAAACAAAGAAAT 1455  
Db 1417 GAGTTTCAACAAGATTAATGTAAGTATGTAAGTAACTGAAAAAGAAATTTAAAGAAAT 1473  
QY 1456 TTCTCAACAGATGATTAATTTATCTGTGTGTAACAGCAGCAAAAGCAGCGGTTGCTT 1515  
Db 1474 TTCCCAACAGATGATTAATTTATTTGTTCTGTGTACAGCAGCAAAAGCTGACGATTACTT 1533  
QY 1516 ACACATGATCTTGAATGAAACATATTTTACGACAAACATCTGATTCAAAAGCGAGAT 1575  
Db 1534 ACACACGATCATCAAAATGAGGCTTACCTTTTACGACAAACATTTGATTCAAAAGAGAT 1593  
QY 1576 TTCCCGACAGATTAATTTGCAAGAAATTTGGGGGAAAGTGGCCAAAGTTTAAAGAGAGAT 1635  
Db 1594 TTTCGACAGATTAAGCTGCAAGAACTAGGAAAAATGCAAGATTTTAAAGAAAGAT 1653  
QY 1636 TCAAAAAGAAATTAATCATGCGCATGTATACACTGAGAGCAATGCAATGCTTGGGACT 1695  
Db 1654 TCAAAAAGAAATTAATCATGCGCATGTATACACTGAGAGCAATGCAATGCTTGGGACT 1713  
QY 1696 ACAGAAATTTTAAAGCAGTTTATTTGATGAACAAAGATTTCTTTGAAAAGACCAAGAA 1755  
Db 1714 ACAGAAATTTTAAAGCAGTTTATTTGATGAACAAAGATTTCTTTGAAAAGACCAAGAA 1773  
QY 1756 AGTTGTGTAATAATCTTAAAGAAAGTGTATTAATGCTGTGAAGAGGAGATGACCGT 1815  
Db 1774 AGTTGTGTAATAATCTTAAAGAAAGTGTATTAATGCTGTGAAGAGGAGATGACCGT 1833  
QY 1816 TTCTCTTTTATGTGTCTTCAAAAACGCTACGTGTGAGCTGATGTTAAAGAGAGTAAA 1875  
Db 1834 TTCTCTTTTATGTGTCTTCAAAAACGCTACGTGTGAGCTGATGTTAAAGAGAGTAAA 1893  
QY 1876 GACAGGTGTAAGTATCAAAAAAATTAATAAGCTTCAATATTAATTAATGATTTCTTGA 1935  
Db 1894 GACAGGTGTAAGTATCAAAAAAATTAATAAGCTTCAATATTAATTAATGATTTCTTGA 1953  
QY 1936 AATTAATCAAAATTAATAAACAACCTGGAAGAAATTTGCTCTTGGCATAAGTATTTGC 1995  
Db 1954 AATTAATCAAAATTAATAAACAACCTGGAAGAAATTTGCTCTTGGCATAAGTATTTGC 2013  
QY 1996 AATGATTTTCACTAATTTTCCAGGTCTTACGAAAGAGATAGTTGTACAAAATCAAG 2055  
Db 2014 AATGATTTTCACTAATTTTCCAGGTCTTACGAAAGAGATAGTTGTACAAAATCAAG 2073  
QY 2056 AAGCATGTGAGCCGTTCTATTAAGAAAGGCTTGAAGATGCTTCAAAAGTGAAGCTT 2115  
Db 2074 AAGCATGTGAGCCGTTCTATTAAGAAAGGCTTGAAGATGCTTCAAAAGTGAAGCTT 2133



Db 1 TGCGCGGCGGCTCAACGCGCAGCAAAAGTGCACAGAAATGACATGATGAGAGCATG 60  
Qy TTTTAGCTTTAATCTTAAAGSAGATGGACTAAGTGAAGCAAGAAATGCAAAAAAATCTAA 121  
Db TTTTAGCTTTGATTTTAAAAAATGATTTGAAGTACAAAATGCAAAATCTAAGTTGG 120  
Qy 122 AAAAATATTCGCAAGAAATGACTGACGCAAACTAAATATAGAACAAATGACACAGAAAC 181  
Db 121 AAGAAATATTCGCAAAACATTAACAAATGACAGATTTAAATCCAGAAAAGTTCCAGAAAAT 180  
Qy 182 TTTAAGTTTTTGGCAAGATGGAAGCAAGATACAAATGCAAAAGAACTGAAAGCCATA 241  
Db 181 TAAAGATTTTCTGTGATACGCGGAAACGAAATGAAAATGTCAAGATCTAATAAACAAG 240  
Qy 242 TTGAGAAAATATGACTACATCAAGGAAAACCTTAAGAGCAATTTAAAAAATATTC 301  
Db 241 TCAATCAAAAATGATTTAAATTCAGGAAAACCTTCAACAGCTGCTAGAAAATAATTT 300  
Qy 302 AGATTATACGGAATGAGATTCAGAAAGATGAAACAACATGCTATTTTGGAGGAG 361  
Db 301 CAGAAATTAACAGATGAGGATTCGAAAAAGATGAACAACATGCTATTTTGGAGGAG 360  
Qy 362 TATGTTCAAAAAGAACTTAAAGATGATTCATCTTGAAGAAATAGTCTATCAAAAAG 421  
Db 361 CATGTCACAGAGAACTTAAAGATGATCAATAAATTAAGGAATTAAGCTTATCAAAAAG 420  
Qy 422 AAGGTGATTAAGTTGCGGAAGAAAGTTCTTTAAGAGCACTTCGTAAGGATCTTAATGAT 481  
Db 421 AACGGAACAATGTGCGAAGAAAGTTCTTTGAGGCGCTTCGTAAGTCTCAATATA 480  
Qy 482 CAGTCATATGTAAGAAAAAATTTAAAGATTTGCGCTGATGCGGAGGAAAGTATG 541  
Db 481 CAAAGACATGTAAGAAAAAATGTAAGAGTTGCCGGAATTTGAAGAAAGAAAGCATG 540  
Qy 542 AGTTAACAACCTGTGTCTGAACCAAGAGACATGAATATTTTAAATGAAAAAG 601  
Db 541 AATTTAAGGAGCTTGTCTTATCAAAAAACAACATGCTTAAGCTTGTAAACAAAAGAA 600  
Qy 602 ATTAAGAGTCCGGAATCTTAAACAGATGTTTACGAGCACTGAGAAAGTTTAA--A 658  
Db 601 AAGTAAATGATGATCTTGAAGAAAGTTGAAGAGCACTTAAGAGAAATTAATTC 660  
Qy 659 AAGAAACATGCTTGAATTACTCGAACATGCTATTTTACATTTGAAATTCGCGAG-- 715  
Db 661 GAAGAAAATGTCTACTTACTTGAAGCAATTTACTTTCACAGAGGAACTGTGAAGAG 720  
Qy 716 -----ACGACGATTAATTAATGTA 736  
Db 721 ACAAAATCAAGTGCAATTAACCTTAATAAAGACTCAAAAGAAATATGTACAGAGTGT 780  
Qy 737 TTTGAATGGAGAGAAATGCCAGAACAAAATTTGCTTATATGCCACAGAGCCGATTT 796  
Db 781 ATGAATTAAGAGAAAGTGTGAAAGAAATAATTTTATATGATCAACAGAGATCGAAT 840  
Qy 797 TTTGATCAACCTAGGACAGAGCTCAATATGAGAGATATATGAGGCTGAAAGATTTTATA 856  
Db 841 TCGATCCAACTAAGCCAGAGCTCACTAGCAGAGGACATGAGGCTGAAAGACTTTATA 900  
Qy 857 AGAAGGTAGAGAGAGATGAGATTTTATTTGAAAGAAATCATCTAAGATGCGACAGCTT 916  
Db 901 AGAGGCGAGAAAGAGATGGAATTTTGTGGAAAGACAAACATGTAAAGATGCAACAGCTT 960  
Qy 917 TGTGGCATTTGTGATCCAGATTTCTAGTCTTAAAAAAGAGAGACAAAGAGAAATGCG 976  
Db 961 TGTGGCACTACT-----TCTTAAGAAAACCTTTAAAAAGAGAAATGTA 1005  
Qy 977 AAGAAAGCTTTCAAAAAGCTGCAAAAATCTCTATGAACATGAGGCTTTAGAAAGTTAT 1036  
Db 1006 TAAAGACCTTTAAABAAAAAATCGCAAAACCTCATGAACATGAGGCTTTAGAAAATCTAT 1065  
Qy 1037 GTAAAGAAAATGTGTTAATGAATATGAGAACGAAAGAAATGTGAAGAAATTTGCAAAATGATA 1096  
Db 1066 GTAAAGAAAATTAACCAAGTGTATGTAAGCAAGAAAATGTGAAGAACTAAGTAAATTAATGATG 1125

Qy 1097 TTAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAAATTAATCGTCTTTTGATCAA 1156  
Db 1126 TTAACAAAACCTTGATCAAGCTTATACATCAACAAATCTTAAABACCGCTTTTACATTTCA 1185  
Qy 1157 CAAAAGAAATTAATGAATTTGTTGATGGAAGGTTGCCAACATTTCTTACACAGAG 1216  
Db 1186 C-----TGATGGAATTTGCCGAAATTTACCGACATTTCTTAATGTGATGAG 1236  
Qy 1217 ATTGTGGAATTTGGAATGCTTATTTGTTTCTAATTTTGAAAAAAATGTCAGATGAGAAA 1276  
Db 1237 ATTGTGCAAACTAGATCTTAATTTGCTTTATTTAATAAGAACTTGTCCAGATTTCAAG 1286  
Qy 1277 ATGATGTAATAAATATTAAGACAAACATGTTACAAAGAGAGACTTATGACAGGCAATA 1336  
Db 1297 AAGCTTGTATGATGTGAGGCGACGCGTTTACAAAGAGAGGCTTGATGACAGGCAAAACA 1356  
Qy 1337 AAGGTGCAAGAAATATGCGAGAAATGTTACATGTTCAACAAAGCTGCTTGA 1386  
Db 1357 GTGTGTCGCAAAAAATATATGCGTGGTTATTAAGTGTCAAAATCAAAAGTTGGCTTAAG 1416  
Qy 1397 AGTTTCAACAAGATTAATTAAGTATGTAAGTAAAGTAAAGAAAGAAAGAAAGT 1456  
Db 1417 AGTTTCAACAAGATTAATTAAGTATGTAAGAGCT--AAAGAAATTAAGAAAGT 1473  
Qy 1457 TCTCAACGATGAAATTTATTTCTGTGTGTAACAGCAGCAAAAGACGCCGTTGCTTA 1516  
Db 1474 TCCCAACGATGAAATTTTGTCTGTGTGTAACAGCAGCAAAAGCTGCAAGATTTACTTA 1533  
Qy 1517 CACATGATCTTGAATGAAACATATCTTTTACGACAAACAACATGATCAAAAGCAGAT 1576  
Db 1534 CACACGATCATCAATGAGGTTACCTTTTACGACAAACAATGAGATCAAAAGAGAT 1593  
Qy 1577 TCCCGACATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCAAAGTTTGAAGAGAT 1636  
Db 1594 TTCGACAGATTAAGACTGCAAGAACTAGGAGAAATCCCAAGATTTGGAAGAGAT 1653  
Qy 1637 CAAAAGAAATTAACATGCGCATGTCATCACTGAGACGACATGCAATCGCTGGGAGCTA 1696  
Db 1654 CAAAAGAAATTAACATGCGCATGTCATCACTGAGACGACATGCAATCGCTGGGAGCTA 1713  
Qy 1697 CAGAAATTTTAAAGCAGGTTTATTTGATGAACAACAAGATCTTTGAAAGCCAGAAA 1756  
Db 1714 CAGAAATTTTAAAGCAGGTTTATTTGATGAACAACAAGATCTTTGAAAGCCAGAAA 1773  
Qy 1757 GTTGTGTAATAATCTTAAAGAAAGTGTAAATGCTCTAAGAGAGATGACCGTT 1816  
Db 1774 GTTGTGTAATAATCTTAAAGAAAGTGTAAATGCTCTAAGAGAGATGACCGTT 1833  
Qy 1817 TCTCTTTGTATGTCCTTCCAAAAGCTACGTTGAGCTGATGTTAAAGACGTGAAG 1876  
Db 1834 TCTCTTTGTATGTCCTTCCAAAAGCTACGTTGAGCTGATGTTAAAGACGTGAAG 1893  
Qy 1877 ACAAGTGTGAAGTATTCAAAAGAAATTAAGGCTTATTAATTAATTTCTTGA 1936  
Db 1894 ACAGGTGTGAAGTATTCAAAAGAAATTAAGGCTTATTAATTAATTTCTTGA 1953  
Qy 1937 ATTAATTAATAAATTAACAACACTGGAAGAAATTTCCCTCTTGGCATACGTAATGCA 1996  
Db 1954 ATTAATTAATAAATTAACAACACTGGAAGAAATTTCCCTCTTGGCATACGTAATGCA 2013  
Qy 1997 ATTAATTTACCTTAATTTGTCAGAGTCTTACGAAGAGATTAATTTAACAATAATCAAGA 2056  
Db 2014 ATTAATTTACCTTAATTTGTCAGAGCTTACGAAGAGATTAATTTAACAATAATCAAGA 2073  
Qy 2057 AGCATGCTGAGCCGTTCTATTAAGAAAGCCCTTGAAGATGCTCTCAAGATGAGCTTC 2116  
Db 2074 AGCATGCTGAGCCGTTCTATTAAGAAAGCCCTTGAAGATGCTCTCAAGATGAGCTTC 2133  
Qy 2117 AAGAAATTTGATGATTAATCTTAATGTGAACCTGATTTGAAGAAATTTGTAAGTATG 2176  
Db 2134 AAGAAATTTGATGATTAATCTTAATGTGAACCTGATTTGAAGAAATTTGTAAGTATG 2193



OY 2177 CGGGAACGTAATATATGCTCAATCACTGCGCTTATGCAAGCTTAACCAAGGTAATCT 2236  
DB 2194 CGGGAAAGTAATAATATGCGTCAATCACTGCGCTTATGCAAGCTTAACCAAGGTAATCT 2253  
OY 2237 CTGGAAGAGTATGATGAGATGCTAGAAAGAACTCTGTGAGAAATTAAGTAAAGAGTGG 2296  
DB 2254 CTGGAAGAGTATGATGAGATGCTAGAAAGAACTCTGTGAGAAATTAAGTAAAGAGTGG 2313  
OY 2297 AAGAACAGTGCAGAAAGCATTTACCAACAGATTAGACAAACCGGACAGCTGATCTAAAAAG 2356  
DB 2314 AAGAACAGTGCAGAAAGCATTTACCAACAGATTAGACAAACCGGACAGCTGATCTAAAAAG 2373  
OY 2357 ATTATTAACATATATGAGAACTTTAAGAAACGTGCGAGAGAGGATGAAACCAAGTCCAGTC 2416  
DB 2374 ATTATTAACATATATGAGAACTTTAAGAAACGTGCGAGAGAGGATGAAACCAAGTCCAGTC 2433  
OY 2417 TTGTTTGTCACTCATTTAAGAAAAACGAAAGTATGATCAAAAAAGTAATAGCAAAAAACA 2476  
DB 2434 TTGTTTGTCACTCATTTAAGAAAAACGAAAGTATGATCAAAAAAGTAATAGCAAAAAACA 2493  
OY 2477 AGGATTAAGATGCGCGTTTCAACCGGACTTCAAGATACCAAAAAACATGTGAAAAATCTAC 2536  
DB 2494 AGGATTAAGATGCGCGTTTCAACCGGACTTCAAGATACCAAAAAACATGTGAAAAATCTAC 2553  
OY 2537 GGAAGGAGGTTAAGATGATATCCGTAAACAGAACTTAAGAACTTAAGCAATTTGATTTGGCAG 2596  
DB 2554 GGAAGGAGGTTAAGATGATATCCGTAAACAGAACTTAAGAACTTAAGCAATTTGATTTGGCAG 2613  
OY 2597 CAGAAATATTGGAAGATATGATGATTTGAAGAAAGATGTAATTAATTTGGAATCAGATT 2656  
DB 2614 CAGAAATATTGGAAGATATGATGATTTGAAGAAAGATGTAATTAATTTGGAATCAGATT 2673  
OY 2657 GCAGAAATTAAGAGAGATTGCAGAAAGCTTGAAGAAAGTATGCAAAAAAGTAAATAGGCTT 2716  
DB 2674 GCAGAAATTAAGAGAGATTGCAGAAAGCTTGAAGAAAGTATGCAAAAAAGTAAATAGGCTT 2733  
OY 2717 GTGCGAATCTGAAGCGCTCTGAGAGGTGAAGCCGCAAGAAACAGTGCAGAAAGTATCAAGA 2776  
DB 2734 GTGCGAATCTGAAGCGCTCTGAGAGGTGAAGCCGCAAGAAACAGTGCAGAAAGTATCAAGA 2793  
OY 2777 CAACATCAACAAACAACAACCGTTGCGGATCCGAGGCAACGGAATGCAAAATCCTTAC 2836  
DB 2794 CAACATCAACAAACAACAACCGTTGCGGATCCGAGGCAACGGAATGCAAAATCCTTAC 2853  
OY 2837 AGACAACAGACACATGGGTTTACACAGACATGACACACACAGACAGCTCTACTATCACAT 2896  
DB 2854 AGACAACAGACACATGGGTTTACACAGACATGACACACACAGACAGCTCTACTATCACAT 2913  
OY 2897 CTACCATCACTCAAAAATTAACATTGACATCAACAGAGCGATGCAAAACCAACCAAGTGT 2956  
DB 2914 CTACCATCACTCAAAAATTAACATTGACATCAACAGAGCGATGCAAAACCAACCAAGTGT 2973  
OY 2957 CGACAGAGAGAGAGATGATGACAGAGACGTGAAACCGAGTGGGGCTGAGATGAGTG 3016  
DB 2974 CGACAGG-----GATGATGTCAGAGAGACGTGAAACCGAGTGGGGCTGAGATGAGTG 3027  
OY 3017 GGTGGAATGTGATGAGGGGGGTGATGATGAGCAATGTTATTTGTTGTTGATTTAG 3072  
DB 3028 GGTGGAATGTGATGAGGGGGGTGATGATGAGCAATGTTATTTGTTGTTGATTTAG 3083

RESULT 7  
AR438436 3090 bp DNA linear PAT 20-FEB-2004  
LOCUS AR438436  
DEFINITION Sequence 5 from patent US 6664053.  
ACCESSION AR438436  
VERSION AR438436.1 GI:42663291  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3090)  
AUTHORS Kovacs,J.A., Huang,S., Masur,H., Fischer,S.H., Gill,V.J. and Mei,Q.

TITLE Identification of a region of the major surface glycoprotein (MSG)  
JOURNAL Gene of human Pneumocystis carinii  
FEATURES Patent: US 6664053-A 5 16-DEC-2003;  
Location/Qualifiers  
SOURCE 1..3090  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 61.9%; Score 1900.6; DB 6; Length 3090;  
Best Local Similarity 77.8%; Pred. No. 0;  
Matches 2430; Conservative 0; Mismatches 609; Indels 84; Gaps 8;  
OY 1 ATGCGCGCGCGGTCAAGCCGCGAGGCGACAGGACACAGAAATACATTTGATGAGAAACAT 60  
DB 1 ATGCGCGCGCGGTCAAGCCGCGAGGCGACAGGACACAGAAATACATTTGATGAGAGCAT 60  
OY 61 GTTTAGCTTTAATTTCTAAGAGAGATGGACTAAGTGCAGAGATGCAAAAAAACTA 120  
DB 61 GTTTAGCTTTGATTTTAAAAAAAAATGATTTAGAAAGTACAAATGCAAAATCTAAGTTG 120  
OY 121 AAAAAATATGTCACAAAGATTGACTGAACCAAACTAATATAGAACAGTACACAGAAA 180  
DB 121 GAAGATATGTCACAAACCTTAAACAAATGCAGATTAATCCAGAAAAAGTTACAGAAAA 180  
OY 181 CTTAAGGTTTTCGGAAGATGAAAGCAGATACAAATATGCAAGAACTGAAAGCCAAAT 240  
DB 181 TTAAGAAATTTCTGTGTAACGGGAAACGAAATATAAATGTCAAGATCTAATAAAACAA 240  
OY 241 ATTGAGAAAAATGTACTCAATCAATCAAGAAAACTTAAGAAAGCAATTAATAAAAAAT 300  
DB 241 GTCAATCAAAATATCATTAATTTTCAAGAAAACTTCAAGACGTGCTGAAAAAAATTT 300  
OY 301 CAGATTAATACGATTAAGATTGCAAGAGATGAAAGATGAACAACATGCTTATTTTGAAGGA 360  
DB 301 TCAGATTAATACGATTAAGATTGCAAGAGATGAAAGATGAACAACATGCTTATTTTGAAGGA 360  
OY 361 GTATGTTCAAAAGAACTTAAGATGATTCATATCTTGAAGAAATTAAGTGTATCAAAAG 420  
DB 361 GCATGTCACACAGAACTTAAGATGATTCATATCTTGAAGAAATTAAGTGTATCAAAAG 420  
OY 421 AAAGGTGTAAGTTGCGGAAAGAGTTCTTTAAGAGACCTTGTAGCGATCTTAATGGA 480  
DB 421 GAACGAAACATGTGGCAAGAGAGTTCTTTGAGGCGCTTGTGTGATCTCAATGAA 480  
OY 481 TCAGTCAATGTGAAAAAACTTAAGAGATTTGCCCTGTATGAGGAGGAAAGTAT 540  
DB 481 ACAAGACATGTGAAAAAACTGAAGAGTTGCCGGAATTAAGAAAGAGGAGAT 540  
OY 541 GAGTTAACAACTTGTGTCTGACACAGAAAGACATGTAAAGATATTTTAATGAAAA 600  
DB 541 GAATTAACGAGCTTTGTCTTTATCAAAAAACAATGCGTAAATGTTGTAACAAAGGA 600  
OY 601 GATAAGAGTGCAGGATCTTTAAGACAGATGTTTACAGACGACTAGAAAGTTTAAA--- 657  
DB 601 AAAGTAAATGTGATCTCTTGAAGAAAGATGTAAGAGACCTTAAGAAAGATGTTG 660  
OY 658 AAAGAAACATGCTCTGAATTAATCTGAAACATGCTAATTTTACATGGAATTCGAG-- 715  
DB 661 CGAGAAAAATGTCTACATTAATCTTGAAGATGTTACTTACAGAGGGAACGTGAAAGGA 720  
OY 716 -----ACGACGATTAATTAATGT 735  
DB 721 GACAAATCAAGTCAATTAACCTAATTAATGAAGCTGCAAGAAATATGATCAAGAGGT 780  
OY 736 ATTGAATTGGAGGAAATGCGCAAGAACAAATATTTGCTTTATATGCGACAGAGCCGAT 795  
DB 781 GATGAATTAAGCAAGAAAGTGTGAAAGAAATATGTTTAATATGATCCAGATCCGAT 840  
OY 796 TTGATCCAACTAAGCCAGAGGCTACATATGACAGAGATATAGGCTGGAAGAGTTTAT 855  
DB 841 TTGATCCAACTAAGCCAGAGGCTACATGACAGAGGACATAGGCTGGAAGAGCTTAT 900

QY 856 AAGAAAGTAGAGGATGAGATTTTATTTGAGAAAGATCATCTAAGAGATGCGACGCT 915  
 DB 901 AAGAGGCGAGAGAGAGATGAGATTTTGTGTGAGAGACAACTGTAAAGATGACACGCT 960  
 QY 916 TTGTTGGCATTTGTGATCCAAAGATTTCTAGTCTTTAAAAAAAAGCGACAAAGAAATGC 975  
 DB 961 TTGTTGGCATCTACT-----TCTTAAGAAAACCTTTAAAAAAGAAAGATGT 1005  
 QY 976 GAAAGAGCCCTTCAAAAAGCTGCAAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTA 1035  
 DB 1006 ATAAAGCCCTTAAAAAAAACCTGCAAAAACCTCATGAAACATGAGGCTTTAGAAATCTTA 1065  
 QY 1036 TGTAAAGAAAAATGCTTTAAGTAATGATGAAAGAAAAATGTGAAGATTCGAATGAT 1095  
 DB 1066 TGTAAAGAAAAATTAACCAAGTAGATGATGAAAGAAAAATGTGATGAATCTGAAGAAAGAT 1125  
 QY 1096 ATTAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATATCGTCTTTTGAATCCA 1155  
 DB 1126 GTTAACAAAACCTTGTCAMGTCTTACATCAACAAATCTTAAAAACCGTCTTTACATTTCA 1185  
 QY 1156 ACAAAAGAAATATGAAATTTGGAATGGAGAGGTTGCCAACATTTCTTAGCAAGAA 1215  
 DB 1186 CC-----TGATGGAATTTGCGAATGGGAAAATTTACCGACATTTCTTAGATGAA 1236  
 QY 1216 GATTGTCGAAATTTGAGTCCCTATTTGTTCTATTGTTGAAAAAAAATGTCAGATGAGAA 1275  
 DB 1237 GATTGTCGAAATTTGAGTCCCTATTTGTTCTATTGTTAAGAAACCTTGTCAGATGTCAAA 1296  
 QY 1276 AATGATGTAAATATTTAAGCAACATGTTTCAAAAAGAGACTTTGATGCGGCAAT 1335  
 DB 1297 GAAGCTGTATGATGAGGCGACCGGTTATTAAGAGAGGCTTGATGCGCGCAAC 1356  
 QY 1336 AAGTGTGCAAGAAATATGCGAGGAATGTTACATGGTTCAAAAGCTGCGCTTGA 1395  
 DB 1357 AGTGTGTCGAAAAATATGCGAGGTTATTTGCTAGCTCAAAATTAAGATGCTTTAAG 1416  
 QY 1396 AAGTTTCAACAGAAATTAAGTAATGAGAAATGTAAGAAATGTAAGAAATGTAAGAAAT 1455  
 DB 1417 AAATTTCAACAGAAATTAAGTAATGAGAAATGTAAGAAATGTAAGAAATGTAAGAAAT 1473  
 QY 1456 TTCTCAACAGTAATTTATTTATCTGTGTGATCAAGCAAGCAAAAGCCGCTTGT 1515  
 DB 1474 TTCTCGAAGCATGATTTGTTGTTCTGTGTATCAACACGCAAGGCGACGATTTCTT 1533  
 QY 1516 ACACATATCTTGAATGAAACATATCTTTTACGACAAAGCATGCAAAAGCGAGAT 1575  
 DB 1534 ACACATCAACATCAAAATGAGGTTATCTTTTACGACAAAGCATGCAAAAGCGAGAT 1593  
 QY 1576 TTCCCGACAGATAAAATTTGCAAGAAATTTGGGAGAAAGTCCAAAGATTTTGAAGAGAT 1635  
 DB 1594 TTTCGCAAGATTAAGACTGCAAGAAATTTAGGAGAAATGCCAAGATTTTGAAGAAAGAT 1653  
 QY 1636 TCAAAAGAAATTCATGCGCATGTCTATACCTGAGACGACAGCAATCGCTTGGGACT 1695  
 DB 1654 TCAAAAGAAATTCATGCGCATGTCTATACCTGAGACGACAGCAATCGCTTGGGACT 1713  
 QY 1696 ACAGAAATTTTAAAGCAGGTTTTTATGAGTGAACACAAAGATCTTTGAAAGACCAAGAA 1755  
 DB 1714 ACAGAAATTTTAAAGCAGATTTTATGAGTGAACACAAAGATCTTTGAAAGATCAAGAA 1773  
 QY 1756 AGTGTGTAAATATCTTAAAGAAAGTGTATTAATGCTTGAAGAGAGATGACGCT 1815  
 DB 1774 AACTGTCCAAAATTTTAAAGAAAGATTTGCAATTAATGCTTGAAGAGAGATGACGCT 1833  
 QY 1816 TTCTCTTTTGTATGTCTTCAAAAAGCTTACGTTGAGCTGATGTTAAAGAGCTGAAA 1875  
 DB 1834 TTCTCTTTTGTATGTCTTCAAAAAGCTTACGTTGAGCTGATGTTAAAGAGCTGAAA 1893  
 QY 1876 GACAGGTGTGAATTTCAAAAAGATTTAAAGCTTCAATATTAATTTGAATTTCTTGA 1935  
 DB 1894 GATAGGTGTGAATTTCAAAAAGATTTAAAGCTTCAATATTAATTTGAATTTCTTGA 1953  
 QY 1936 AATAATATCAAAATTAATAACCAACCTGGAAGAAAGAAATGTCCCTCTTGCAATACGATTTGC 1995

---

DB 1954 AAAAATCAAAATTAAGCAGAAATCAGCAGCAAAATTTGTCCCTCATGCGCATCTACTGC 2013  
 QY 1996 AATGATTTTCACTAATTTGTCCAGGCTTACGAAAGAGATGT---TGTACAAAATC 2052  
 DB 2014 GATGATTTTACCAATTTGTCCGATCTTAAAGAGAAAGAAATTTTGTGTCAAAATCTT 2073  
 QY 2053 AAGAGCATGTGAGCCGTTCTAATTAAGAAAGCCCTTGGAAGATGCTTCAAGTAGAG 2112  
 DB 2074 AAAAAATTTGCGAACCATTTCTACAAAAGAGTTTGAAGATGCTTTAAAGTAGAG 2133  
 QY 2113 CTTCAAGAAAATTTGACTGATTAATCTTAAATGTGAACCTTGATTTGAAAGATTTGTA 2172  
 DB 2134 CTTCAAGAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2193  
 QY 2173 GTAGCGGAAACGTAATTAATGCGTCAATGAGGCTTATGCAAGCTTACCAAGAGAT 2232  
 DB 2194 GTATTTGAAGCGTAATTAATGCGTCAATGAGGCTTATGTAAGATTAATTAATTAATTA 2253  
 QY 2233 AACTCTGAAAAG---AGTGAATGAGATGCTAAGAAAGACTGTGTGAATAATAGTAA 2289  
 DB 2254 AAAAAATTAAGCGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2313  
 QY 2290 GAAGTGAAGACAGTGCACAAAGCATTTACCAACAGATTTAGACAAACCGGACGCTGATCTA 2349  
 DB 2314 GAGGTGAAGACAGTGCACAAAGCATTTACCAACAGATTTACCAACAGCTGCAAAAAAGCTTA 2373  
 QY 2350 AAAAAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2409  
 DB 2374 AAAAAATGTTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2433  
 QY 2410 TCCAGTCTTGTGTTGTCACTCATTAAGAAAAAGAAAGTAATGTCAAAAAAGTAATAGC 2469  
 DB 2434 TCCAGCTTGTGTTGTCACTCATTAAGAAAAAGAAAGTAATGTCAAAAAAGTAATAGC 2493  
 QY 2470 AAAAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2529  
 DB 2494 AAAAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2553  
 QY 2530 ATACTAGGAGGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2589  
 DB 2554 ATACTAGGAGGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2613  
 QY 2590 TTGGCAGCAGAAATTTTGAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2649  
 DB 2614 TTGGCAGCAGAAATTTTGAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2673  
 QY 2650 TCAAGATTTGCAAGATTAAGAGATTTGCAAGAACTTGAAGAAATGATCAAAAAAGATTAAT 2709  
 DB 2674 TCGGATTTGCGGAAATTAAGAGATTTGCAAGAACTTGAAGAAATGATGGAAGATTTGAG 2733  
 QY 2710 AAGCTTTGTGCAATCTGAAGGCTCTTGAGAGTGAAGCCGACGAAACAGTGAAGAAAGT 2769  
 DB 2734 AAGCATGTCAAGATTTGAAGGCTCTTGAGAGTGAAGGATGATGAAGAAATGTCACGAAAGC 2793  
 QY 2770 ACAACGACAACTCAACAAACAAACAAACCGTTGCGATCGAAGGCAACGGAATGCAAA 2829  
 DB 2794 ACAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2853  
 QY 2830 TCTTACAGACAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2889  
 DB 2854 TCTTACAGACAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2913  
 QY 2890 ATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2949  
 DB 2914 ATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2973  
 QY 2950 AAGTGTACGACAGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3009  
 DB 2974 AAGTGTACGACAGG-----GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3027  
 QY 3010 ATGAGGTGTGAATGTGATGAGGAGGAGATGATGATGATGATGATGATGATGATGATGATGAT 3069

Db 3028 ATGATGCGCTGAGCGGTATGAGGGGGGTATGATGACATGTTATTTCCTGATGATT 3087  
Qy 3070 TAG 3072  
3088 TAG 3090

RESULT 8  
AF033208 3089 bp DNA linear PLN 10-SEP-1998  
LOCUS Pseudocystis carinii f. sp. hominis clone HUMSG11 major surface  
DEFINITION glycoprotein (MSG) gene, partial cds.  
ACCESSION AF033208  
VERSION AF033208.1 GI:3560512  
KEYWORDS  
SOURCE Pseudocystis jirovecii  
Pseudocystis jirovecii  
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 3089)  
Garbe, T.R. and Striinger, J.R.  
TITLE Molecular characterization of clustered variants of genes encoding  
major surface antigens of human Pseudocystis carinii  
JOURNAL Infect. Immun. 62 (8), 3092-3101 (1994)  
MEDLINE 94314421  
PUBMED 7518806  
REFERENCE 2 (bases 1 to 3089)  
Mei, Q., Turner, R.E., Social, V., Livingston, D., Angus, C.W. and  
Kovacs, J.A.  
TITLE Characterization of major surface glycoprotein genes of human  
Pneumocystis carinii and high-level expression of a conserved  
region  
JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)  
MEDLINE 98380374  
PUBMED 9712777  
REFERENCE 3 (bases 1 to 3089)  
Mei, Q., Turner, R., Social, V., Livingston, D., Angus, C.W. and  
Kovacs, J.A.  
TITLE Direct Submision  
JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National  
Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD  
20892-1662, USA  
FEATURES  
source  
1..3089  
/organism="Pseudocystis jirovecii"  
/mol\_type="genomic DNA"  
/specific\_host="Homo sapiens"  
/sub\_species="hominis"  
/db\_xref="taxon:42068"  
/clone="HUMSG11"  
/note="derived from HIV-infected human with P. carinii  
pneumonia"  
1..3089  
/gene="MSG"  
/note="surface antigen"  
/codon\_start=3  
/product="major surface glycoprotein"  
/protein\_id="AAC34971.1"  
/db\_xref="GI:3560513"  
translation="ARAVRRRAKGAQNSIDIEHVALILKNGLEDYKCTKLBEYCK  
LTNAGLNPKEVHEKLDKCDNGKNEKODLKNVNGKICFKQGLQTAAGKISBL  
TDECKKEBOCLFLEGACPTLKDCKNLKNNCYOKERNVVAEVLALRGDLNET  
KTCEKKKEVCKLRBSDELTELCLYQKTTCVSLVTGKSKCDTLEKEVEBALKNE  
LREKLLIBOCYFHRNCGEKGSKCNPNKDCKEVYPECEBLAKGKKNIVYMH  
GSDPPTPEPTLADIGUBELYKRAEDGIFVGRQVRDATALALILKTLKKEBC  
1 KALKNCENPEHEBALNLCKENRPSDGTGKCDLKVKTCTSLTSTLTKRLY  
ISPDIAGWGLPTFLSDDCALKESYCFYVETCPDYVEACMNTAACYRGDLARA  
NSVLQNNKGLGSKNDMLKKFQOELAVCEKLGKNGSFENDELFLVLCIOPAKAR  
LITTHQWRVIFLRQLOKRPDPTDCKELGRKCDLGRGDSKEITTPCHTLBOCN  
RLGITREILKQILIDEHKDTLKSHEKCAQITLAKCKHKRBRGDRSPFCVQNAETLK  
NVKDVQDRCKLFEENMQASDINDSLKQIKAESANICPSWHPYCDRFLPNCPLRK

## ORIGIN

Query Match 61.8%; Score 1899.6; DB 8; Length 3089;  
Best Local Similarity 77.8%; Pred. No. 0;  
Matches 2429; Conservative 0; Mismatches 609; Indels 84; Gaps 8;  
GKTFQNLKXKCEPFYRKYVLEDAIKYELKGNLSNITKCEPALERYCTVLKDVNNA  
SLICKDNTESTKTKADKNVRKKLCLLVEBEQOCVLPTELTELESKLDVKTBE  
ELKERAKKANNSKSLVLKVNESNTSKNSKXNDKRVNSGQDITKYKILRGV  
KEALVTESEAKAFDLAEVFGRYVDLKEKCEKITSDCGIXDDCGDLKEVCKIEKTH  
DLKPLEYKSHIETSTTTTTTTTVDLPKATECKSLQDTTDVTQSTHTSTSTIT  
STISKTITLSTRCKPTKCTTGDEADVPRSEBLKXNSGMSVMGVIIVAMVISPMI"  
2 TGGCGCGGCGGCTGACGCGGCGAGCAGGACACAGATAGCTTATGAGAACATG 61  
1 TGGCGCGGCGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
62 TTTTATGCTTTAATCTTAAAGAGATGAGACTTAAAGAGAGAGAGAGAGAGAG 121  
TTTTATGCTTTAATCTTAAAGAGATGAGACTTAAAGAGAGAGAGAGAGAGAG 120  
61 TTTTATGCTTTAATCTTAAAGAGATGAGACTTAAAGAGAGAGAGAGAGAGAG 120  
122 AAAAATATGCGAAGAAATTTGACTGAGCAAACTTAAATATAGAAACAATGACA 181  
121 AAGATATGCGAAGAAATTTGACTGAGCAAACTTAAATATGAGAAAGTTGACA 180  
182 TTTAAAGTTTTTGGCGAAGATGAGAAAGCAGATACAAATGCAAGAACTGAAAG 241  
181 TTTAAAGTTTTTGGCGAAGATGAGAAAGCAGATACAAATGCAAGAACTGAAAG 240  
242 TTTGAGAAAAATGTAATCTTAAAGAGAAATTTAAAGAGCAATTTAAAAAATTC 301  
241 TCAATCAAAAATGCAATTTAAATTTTCAAGGAAATCTTCAACAGCTGCGAAAAATTT 300  
302 AGATTATTAACGATTAAGATTTGCAAGAAAGAAATGCAAACTTATTTTGAAGGAG 361  
301 CAGATTATTAACGATTAAGATTTGCAAGAAAGAAATGCAAACTTATTTTGAAGGAG 360  
362 TATGTTCAAAAGAACTTAAAGATGATTTGCAATCTTGAAGAAATTAAGTCTATCA 421  
361 CATGTCCAAAGAACTTAAAGATGATTTGCAATCTTGAAGAAATTAAGTCTATCA 420  
422 AACGTGATTAAGTTTGGCGAAGATGCTTTTAAAGAGCACTTGAAGCATTTAATGAT 481  
421 AACGATTAAGTTTGGCGAAGATGCTTTTGAAGGCGCTTCTGATGATCTCAATGAA 480  
482 CAGTCATATGTAAGAAAAAATTTAAAGAGATTTGCTGTCATGCGGAGGAGAAAGTGA 541  
481 CAAAGATATGTAAGAAAAAATTTAAAGAGATTTGCTGTCATGCGGAGGAGAAAGTGA 540  
542 AGTTTAAAGAACTTGTGCTGTAACCAAGAAAGACATGTAAGAAATTTTAAATGAA 601  
541 AATTAAAGCAAGCTTGTGCTTAAAGAAATTTAAAGAAATTTTAAAGAAATTTTAA 600  
602 AATAAGAGTGGGTAATCTTAAAGCAAGATTTTGAAGGAGCACTGAGAACTTTTAA 658  
601 AAGTAAATGTAATCTTAAAGCAAGATTTTGAAGGAGCACTTAAAGAAATTTTAA 660  
659 AAGAAACATGCTTGAATTTTAAAGCAAGATTTTAAAGGAGGAGGAGGAGGAGGAG 715  
661 GAGAAAGATGCTTAAATTTTAAAGCAAGATTTTAAAGGAGGAGGAGGAGGAGGAG 720  
716 -----ACGAGATTAATTAATGTA 736  
721 ACAAATCAAGATGCAATTAACCTTAATTAATTAAGCACTGCAAGAAATTTTAA 780  
737 TTGAATGGGAGGAGAAATGCAAGAAATTTTAAATGCACTGAGGAGGAGGAGGAG 796  
781 ATGAATTAAGCAAGAAATGCAAGAAATTTTAAATGCACTGAGGAGGAGGAGGAG 840  
797 TTGAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 856  
841 TGATTCACCAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
857 AGAAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 916



Db 901 AGAGGGCAGAGAGAGATGGAATTTTGTGGAGACACATGATAGAGATGCACAGCTT 960  
Qy 917 TGTGGCATTTGTGATCCAGATTTCTAGCTTTAAAAAAGAGACAAAGAAATGCG 976  
Db 961 TGTGGCATCTACT-----TCTTAAAGAAACCTTTAAAAAAGAAATGTA 1005  
Qy 977 AAGAAGCCCTTCAAAAAAGCTGCACAAAATCTCATGAACATGAGGCTTAGAAAGTTAT 1036  
Db 1006 TAAAGCCCTTAAAAAAGCTGCAGAAACCTCATGAACATGAGGCTTAGAAATCTAT 1065  
Qy 1037 GTAAAGAAAATGTTTAAAGTAAATGAGACGAAAAATGTGAAGAAATGCAAAATGATA 1096  
Db 1066 GTAAAGAAAATTAACCAAGTATGATGACGAAAAATGTGATGACTAGAAAAAGATG 1125  
Qy 1097 TTAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTTTTGTATCCA 1156  
Db 1126 TTAACAAAACCTTGTAACAAGTCTTACATCAACAAATCTTAAAAACGCTCTTACATTTTCAC 1185  
Qy 1157 CAAAAGGAAATTAATGAATTTGTGAATGGGAAGGTTGGCAACATTTCTTAGCAACGAG 1216  
Db 1186 C-----TGAATGAATTTGGCAATGGGAAATTAACCAATTTCTTAGTGAAG 1236  
Qy 1217 ATTTGCGAAATTTGAGTCTTATTTCTATTTTGAAGAAAAATGTCAGATGAGAA 1276  
Db 1237 ATTTGCGAAATCTGAATCTTATTTCTTATTAAGAAACCTTGTCCAGATGTCAAG 1296  
Qy 1277 ATGATGTAAATATTAAGACACATTTACAAAAGAGACTTGATGACCGGCAATA 1336  
Db 1297 AAGCTTGTATTAATGTGAGGCGAGCGTGTATTAAGAGAGGCTTGATGACGGGCAACA 1356  
Qy 1337 AAGTGTGCAAGAAATTAATGAGGAGATGTTACATGTTCAAAACAAAAGCTGGCTGAA 1396  
Db 1357 GTGTTGTCAAAAAATATGAGAGGCTTATGCAAGCTCAAAATTAAGATTTGCTTAAAG 1416  
Qy 1397 AGTTTCAACAAGATTTAGTAAAGATATGTGAAACCTGAAAAAGAAACAAAGAAAT 1456  
Db 1417 AATTTCAACAAGATTTAGCAAAAGTATGTGAAGAACTG---AAGGAAATTAAGAAAT 1473  
Qy 1457 TCTCAACGATGAATTTATTTATCTGTGTGTAACGCCGCAAAAGACCGCGTTGCTTA 1516  
Db 1474 TCTGAAACGATGAATTTGTTTCTGTGTATTAACAACGAAAGGCAAGCAAGATTACTTA 1533  
Qy 1517 CACATGATCTTGAATGAAGAACTATCTTTTACGACAACTGGATCAAAAGGAGAT 1576  
Db 1534 CACATCACATCAATGAGATTTCTTTTACGACAACTGGATCAAAAGGAGAT 1593  
Qy 1577 TCCGACAGATTAATAATGCAAGAAATGGGAGAAAGTGCCAAGATTTAGAGAGAT 1636  
Db 1594 TCCGACAGATTAAGATCTGCAAGAAATAGGAGAAATGCCAAGATTTAGAAAAAGAT 1653  
Qy 1637 CAAAAGAAATTAATGATGCGCATGTCACTAGAGCAGCAATGCATGCTGGGAGTA 1696  
Db 1654 CAAAAGAAATTAATGATGCGCATGTCACTAGAAACAGCAATGCAATGCTTAAAGGATTA 1713  
Qy 1697 CAGAAATTTTAAACAGGTTTTTATTTGATGAACACAAAGATCTTTGAACCAAGAA 1756  
Db 1714 CAGAAATTTTAAACAGATTTTATTTGATGAACAAAGATCACTTTGAAAAGTCAAGAA 1773  
Qy 1757 GTTGTGTAAATACCTAATAAAGAAAGTAAATGCTTAGAAGAGAGATACCGTT 1816  
Db 1774 ACTGTGCAAAATATTTAAAGAAAGAAATGCCATTAATGCTTAGAAGGATGATGCTT 1833  
Qy 1817 TCTCTTTTGTATGTGTCTTCCAAAACGTAAGCTGTAGCTGATGTAAGAAGCTGAAG 1876  
Db 1834 TTTCTTTTGTATGTGTCTTCCAAAACGTAATGAGCTAGCTGATGTAAGAAGCTGAAG 1893  
Qy 1877 ACAGGTGAAGTATTCAAAAGAAATTAAGAGCTTCAATATTAATGAATTTCTGAAA 1936  
Db 1894 ATAGGTGCAAAATATTTCAAGAAAGAAATGCAAGCATGATTAATGATTTCCCTTAAA 1953  
Qy 1937 ATAAATACAAATTAATAACAACATGCAAGAAAGATTTGCTTGTGATACGTATTTGCA 1996

Db 1954 AAAATCAATTAAGCAAAATCAGCAGCAAAATATTTGCTCATGCAATCCATCTCGC 2013  
Qy 1997 ATAGATTTTACCAATATTTGTCAGGCTTACGAAGAGAAATAGT---TGTCAAAAAATCA 2053  
Db 2014 ATAGATTTTTCACCAATATTTGTCAGATCTTAAGAAAGAAACCTTCTGTCAAAATCTTA 2073  
Qy 2054 AGAAGCATGTGAGCCGCTTCTAATAAAGAAAGGCTTGAAGATGCTCTCAAGATAGNC 2113  
Db 2074 AAAAATATTTGGAACCAATTTCTACAAAAGAAAGGTTTAGAAGATGCTTTAAAGTAGC 2133  
Qy 2114 TTCAAGAAAATGACTGATTAATCTTAATGTGAACCTGCATTTGAAGAAATATTTGACG 2173  
Db 2134 TTTCAAGAAAATTTAAGTAAATATACTTAATGTGAACCTGCATTTGAAGAAATATTTGACG 2193  
Qy 2174 TAGGGGAAAACGTAAATTAAGCCGCAATCGATGCTTATGCAAGCTTAACCAAGGATA 2233  
Db 2194 TATTGAAGACGTAAATTAATGCGTCAATCAGCACTTATGTAAAGATTAATCCAAAGTA 2253  
Qy 2234 ACTCTGAAAAG---AGTATGAGGATGCTAGAAAGAACTGTGAGAAATTTAGTGAAG 2290  
Db 2254 AAATCTAAAAAGCCGATTAATTAATTTAGAAAGAAAGCTTTGTCTAATATTTAGTGAAG 2313  
Qy 2291 AAGTGAAGAACAGTGCAGAACATTTACCAACAGAAATTAAGACAAACCGCAGCTGATCTAA 2350  
Db 2314 AGGTGAACAGCAATGCAGAAAGTATTAACCAACAGAAATTAACAGAGCTGAAAAAGTCTAA 2373  
Qy 2351 AAAAAGATTAATAGACATTAAGAGAACTTTAAGAAACGTGCAGAGAGAACATGAACAT 2410  
Db 2374 AAAAAGATTTTAAACATTAAGAGAACTTTAAGAAAGGCAAAAAAGCAATGAACAGT 2433  
Qy 2411 CCACTTTGTTTTTCACTCACTTAAGAAAAAGAAATATATGTAATGTAATGCA 2470  
Db 2434 CCACTTTGTTTTTCACTTTGTTAGAAAAAGAAATATATGTAATGTAATGCA 2493  
Qy 2471 AAAACAAAGATTAAGATGCGCTTTCAACCGCACTTCAAGATTAACCAAAACATGTGAAA 2530  
Db 2494 AAAACAAAGATTAAGATGCGCTTTCAACCGCACTTCAAGATTAACCAAAACATGTGAAA 2553  
Qy 2531 TACTACGAGAGGAGTTAAGATGATCCGTTAACAGAAATTAAGAGCTTAAGCATTTGATT 2590  
Db 2554 TACTACGAGAGGAGTTAAGAGGCACTTGTAAACAGATCTGAAGCCAAAGCATTTGATT 2613  
Qy 2591 TGGCAGCAGAAATTTTGAAGATATGATGATTTGAAGAAAGATGTAATTAATTTGAAAT 2650  
Db 2614 TGGCAGCAGAAATTTTGAAGATATGATGATTTGAAGAAAGATGTAATTTGAAAT 2673  
Qy 2651 CAGATTCAGAAATTAAGAGAGATTTCAAGACTTGAAGAAATGTAATTAATTAATA 2710  
Db 2674 CGATTTGCGGGAATTAAGAGAGATTTGCGATGCTTGAAGAAATGTAATTAATTAATA 2733  
Qy 2711 AAGCTTTGCGCAATCTGAAGCTTCTGAAGGTGAAGCCGACAGAAACATGACAGAAAGTA 2770  
Db 2734 AGACATGTCAAGATCTGAAGCTTCTGAAGGTGAAGTGCATGAAGATGTCAACAGAAAGCA 2793  
Qy 2771 CAAACGAACTTAAGCAACAAACAAACCGTTGCGCATCCGAAGGCAACGGAATGCAAT 2830  
Db 2794 CAAACGAACTTAAGCAACAAACAAACCGTTACCAATCCGAAGGCAACAGAAATGCAAT 2853  
Qy 2831 CTTTACAGACAAACAGACATGAGTTTACAGACATTCAGACACACAAACAGCTCTACTA 2890  
Db 2854 CTTTACAGACAAACAGATTAATGAGTTTACAGACATTCAGACACACAAACAGCTCTACCA 2913  
Qy 2891 TCAATCTTACATTAATCAAAATTAACATTTGACATCAAGAGGCGATGCAACCAACCA 2950  
Db 2914 TCAATCTTACATTAATCAAAATTAACATTTGACATCAAGAGGCGATGCAACCAACCA 2973  
Qy 2951 AGGTACAGACAGAGAGAAAGATGATCAGAGAGCGTGAACCGAGTGAAGGCGCTGAGA 3010  
Db 2974 AGGTACAGACAGG-----GATTAAGCAGAGAGCGTGAACCGAGTGAAGGCGCTGAGA 3027  
Qy 3011 TGAGTGGGTGAATGATGAGGGGGGTGAATGATGCAATGTTATTTGTTCAATGATTT 3070  
Db 3028 TGAGTGGGTGAAGCTGATGAGGGGGGTGAATGATGCAATGTTATTTGTTCAATGATTT 3087

RESULT 9  
 LOCUS PNCMSGI 3363 bp DNA linear PLN 26-SEP-1994  
 DEFINITION Pneumocystis carinii B-cell receptor (megi) gene, 3' end.  
 ACCESSION L27092  
 VERSION L27092.1 GI:535706  
 KEYWORDS B-cell receptor.  
 SOURCE Pneumocystis carinii  
 ORGANISM Pneumocystis carinii  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystis carinii.  
 REFERENCE 1 (bases 1 to 3363)  
 AUTHORS Garbe,T.R. and Stringer,J.R.  
 TITLE Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii  
 JOURNAL Infect. Immun. 62 (8), 3092-3101 (1994)  
 MEDLINE 94314421  
 PUBMED 7518806  
 COMMENT Original  
 FEATURES  
 source source text: Pneumocystis carinii DNA.  
 location/Qualifiers  
 1..3363  
 /organism="Pneumocystis carinii"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4754"  
 152..3244  
 /gene="megi"  
 <152..3244  
 /gene="megi"  
 /codon\_start=1  
 /product="B-cell receptor"  
 /protein\_id="AAA21645.1"  
 /db\_xref="GI:535707"  
 /translation="VARAAKQVAVGVKNNAEBERFLALIMRADYKDSCKNKKIKYCV  
 DGIKRNLTSEBVKHEIKEDPFCKDSGSGKKEELKKNVAKCANNPEPTLEGIYKVDASG  
 LTNDCKENBEOCLPLEBACEDPLVEDGSKLNRCLCYQKRBGVAEYVLLRALRGLGNK  
 TECEKRIKDVCPRKIGQSDDELTLCLDQKCTTNLMTRDKRCNTLEEDYKALENKRN  
 NLGKCLPLBHAFTTEGTAAYKASQCTENKQCELDLPKCDLAEBCGKKGIYIHGPG  
 DPFLEPPEPTAAEDIGLELYAKAEDSVNHCPEVPVATALLALLIONPPKIOANE  
 KKKCVLENKCKELKHEIVSDGLCNQAAASGFGKCEELKEELANSTKLISEKIKR  
 KHLGSGEITTPWYLTSTPLSDPSDARLESDFEPADODPLKCKCKVAKACVGRGL  
 ARAKNTLOENRGLLRGNSQWLVKRFQELKYVCEKLEENKGSFSDNDELVYLCVPRP  
 KAARLTHDLRWTTIFLRQQLDQKDFPTVTAKIRKCODLGGFQKEITWPCHTT  
 EQQCRLGLTTEIILKQVLLNHEKDTLKTENECVYLKKECNMSRGRGDRFSFVGFON  
 ATCKLMVDOVDRCKIPEKNIKVEIIVDFLKNNTNITTELRNCPSHWTYCNRPSSNQ  
 PDFSKNPECTIKNNCKEPEERKALEDLKVELRGLKSDENKCTPALKGYCTLAGNVAN  
 NASVRSICKDNQNSNKKTDKVVSELCKKIMEVYKQCCETLPALRQPADLDEKDVV  
 TYBELKEAKAKAMKSSLYLVFVKKDGNNTRKNNKSDKRVVSEKDTTIGVYKILRR  
 GVKDVLTVELAKAFDLAELVFGRYVDLKECEKTLUDCGIKDDCDGLKVGCGKIKRR  
 CRDLKPLEVKSHEIVTESITTTTTTTTTTTTTVTDPKATCEKSLQTTDTWVQTSHTS  
 ITSTTSKTLITLSTRCKPRTCTTGDADVDVPSGGLRVSGMNWRGAILVAVISFMI  
 152..3241  
 /gene="megi"  
 /product="B-cell receptor"  
 /standard\_name="major surface antigen"  
 /evidence=experimental

Db	213	TTTTTGGCTTGGATTATATAGAGCAGATTATTAAGATGAAAGTAAATGTGTAATAATTAATA	272
Oy	122	AAAAATATTCGCAAGATTTGACTGAAAGCAAACTTAATATATGAAACAATGACACAGAAAC	181
Db	273	AMGAATTTTCGATGGGTTAAAAAATGCTTCACTAACATGAGAAAGTACTTAAGAT	332
Oy	182	TTAAAGTTTTTGGCAAGATGGAAGAAAGCAGATACAAATGCAAAAGACTGAAAGCCAAATA	241
Db	333	TAAAGATTTTTTGCAAAAGATGGAAGTCAGAGAAAATAATGCGAAGAACTAAAAAAAATG	392
Oy	242	TTTGGAATAAATGTACTACAACTCAAGAGAACTTAAGAAAGCAATTAATAAAAAAATTC	301
Db	393	TTGGAAGCAAAATGCAATTAATTTTAAACAAAATTTGAAAGATGGTAAAAAAGATGCTT	452
Oy	302	AGATTATTAACGGATTAAGATTTGCAAAAGAAATGAAACAACATGCTTATTTTGGAGGAG	361
Db	453	CAGCTTGACAAATGATGATTTGTAAAGAAATGAACGACAAATGCTTATTTTGGAGGAG	512
Oy	362	TATGTTCAAAAGACTTAAGATGATTCAAATCTTTAGAAATTAAGTACTATCAAAAGA	421
Db	513	CATGCCCTTGTAGATGAGATTTGTAGCAAACTTAAGGAATCTATGTTTACAGAAAA	569
Oy	422	AACCTGATTAAGTTGGCGGAAGAGTTCTTTTAAGAGACTTCGTAAGGATCTTAATGAT	481
Db	570	AACGTGAAGGATGACGAAGAAAGTCTTTTGAAGGACATTCGTGGGATTTTGAATA	629
Oy	482	CAGCTATATGTAATAAATAAATTAAAGATTTGGCCCTGATCATGGGAGGGAAGTATG	541
Db	630	AAACTGAAATGCAAAAGAAATTAAGATGTTTGGCCAAAATATGCGCAAGAAAGTATG	689
Oy	542	AGTTAACAACTTGTTGTTCTGAACCAAGAAAGACATGTAAAGATATTTTAATTGAAAAAG	601
Db	690	AGTTGACATTTGTTGTCTTGATCAGAAAGAAACATCAGAACTTAATGACTGCGGAG	749
Oy	602	ATAAGAATGCGGACTCTTAAACAAATGTTTCAGAGACATTAAGAAATTTTAAAAA--	659
Db	750	ACAAAAGTGAATACCTTGAGGAAGATGTTAAAAAGCATGGAAGCAAAAACAACT	809
Oy	660	----AGAAACATGCTTGAATTACTCGAACATGCTATTTTTCATTTGAAATTG-----	710
Db	810	TGTTAGAAATATGTTTGCCATTTACTTGAGC-AATGCTATTTTCAAGAGGGGACATGCAAA	868
Oy	711	-----CGAGAGCAGATATTAATTAATGTAATTGAA	741
Db	869	AAACATCAAGTGCACCTCCAAATTAAGACTGCGAGAGCTATCTACCAAAAGTGTGATGAA	928
Oy	742	TTTGGAGAAAAATGCCAAGAACAAATATTTGCTTATATGCCACGAGACCCGATTTTGAT	801
Db	929	TTGGCAAGAAATGTGAAAAAAGGGCATATTTACATACATCCGGAGCCCGATTTTGAT	988
Oy	802	CCAATGAGCCAGAGGCTACAAATGACAGAGATTTAAGGCTGGAAGATTTTATTAAGAG	861
Db	989	CCAATGAGCCAGAACTTACAGTACAGAGACATAGGCTGGAAGCTTTTATTAAGAG	1048
Oy	862	GTAGAGAGATGAGATTTTATTTGAAGAAATCATTAAGAGATGCGACGCTTTGTG	921
Db	1049	GCCGACAGAAATGTGTTCTATATCGAAAGGCTCTGTAAAGAGCGCAACTGCTCTACTG	1108
Oy	922	GCATTTGTGATCCAAAGTTCTAGCTTTAAAAAAAAG---ACGACAAAGAAATGCGAA	978
Db	1109	GCACCTTTGATTCGAATCCGAATCCGAATTCACACGAAAGCAAAAAAGAAAAATGCAAA	1168
Oy	979	GAAAGCCCTTCAAAAAGCTGCAAAATCTCTCATGAACATGAGGCTTTAGAAAGTTTATGT	1038
Db	1169	AAAGTTCTTGAAAAATAGTGTAAAGATTAATAAAAAACATGAGTTTGGAGATCTTTGT	1228
Oy	1039	AAGAAAAATGTTTAAGTAATGATGGAACGAAAAAATGTGAAGATTTGCAAAATGATATT	1098
Db	1229	AATCAAAATGACGTAATGCAAAAGTGAACCTAAAAAGTGTGAAGAACTAGAAAGAAATTA	1288
Oy	1099	AACAAAATTGCAAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTGATCCACA	1158

Db	1289	GCAGAAACGTA	CTAAAAATCTTTCTGTGAAAAAATAAGAAATAAACATCTCTCT	-----	1339	
Qy	1159	AAGGAAATATATGA	AAATTTGTTGGATGGGAAGGTTGCCAACATTTCTTAGCAAGAAAT	1218		
Db	1340	---	-GGATCCGAGAGAAACCATTCATCGTATTAAGTTGTGCAATTTCTTAGGACAGTGAC	1396		
Qy	1219	TGTGCGAAAT	TGGAGTCCCTATTGTTCTATTTTGAAAAAATGTCAGATGAGAAAT	1278		
Db	1397	TGTCCAGGTTAGAT	CGACTGTTTTTATTTTCTCAAGATTAAGATCCTTTAAAAA	1456		
Qy	1279	GCATGTAAAAATAT	TAAAGACAACTGTATCAAAAAGAGACTTTGATGACGGGCAATATA	1338		
Db	1457	GAATGCAAGAT	TGTTAAGCAGCACTGTTATTAAGAGGGCTTGATGACGGGCAATATA	1516		
Qy	1339	GTCGTGCAGAAAA	TATGCGAGAAATGTTACATGCTTCAACCAAAAGCTGGCTTGAAG	1398		
Db	1517	GTCGTGCAGAAAA	TATGCTGGTTATTA	CGTGGTTCAATCAAAAGTTGGCTTAAGAA	1576	
Qy	1399	TTTTCAACAGAA	TTTAGTAAAGTATGTGAGAACTGAAAAAAGAAACAAAGAAATTC	1458		
Db	1577	TTTTCAACAGAA	TTTAGTAAAGTATGTGAGAACTGAAAAAAGAAATTAAGAAATTTT	1636		
Qy	1459	TCAAAAGCATGA	TAATTTATTTCTGTGTGTACAGCGACAAAGAGCCGGTTGCTTACA	1518		
Db	1637	TCAAAAGCATGA	TAATTTGTTCTGTGTGTACAGCGACAAAGAGCCGGTTGCTTACA	1636		
Qy	1519	CATGATCTT	CGAATGAAAACTATCTTTTACGACAAACATCGATCAAAACGAGATTC	1578		
Db	1697	CATGATCTT	CGAATGACATGACAACTATCTTTTACGACAACTGGATCAAAACGAGATTTT	1756		
Qy	1579	CCGACAGAT	AAAAATTGCGAGAAATTTGGGGAGAAAGTCCCAAGTTTAGAGAGATT--	1636		
Db	1757	CCGACAGAT	AAAAAGACTGCGCAAGGAAATTAAGGAGAAATGCCCAAGTTTAGGAAAGGATTC	1816		
Qy	1637	-	CAAAAGAAATTACATGGCCCATGTATATCACTGGAGCAGCAATCGCTTGGGGACT	1695		
Db	1817	CAAAAAGAAAT	TACGTGGCCATGTATATCACTGGAGCAGCAATCGCTTGGGGACT	1876		
Qy	1696	ACAGAAATTTT	AAAGCAGGTTTATTTGGATGAACAAGATTA	CTTTGAAAGACCAAGAA	1755	
Db	1877	ACAGAAATTTT	GAAGCAGGTTTATCTAATGAACAAGATTA	CTTTGAAAGACCAAGAA	1936	
Qy	1756	AGTTGTGTA	AAATATCTAATTAATGCTTAGAAGAGAGATGACGT	1815		
Db	1937	AACGTGTAA	CATTTTAAAGAAAGGTAAATTAATGCTTAAGAGGGCGAGATGACGT	1996		
Qy	1816	TTCTCTTT	TGATATGCTTCCAAAGCCGTACGTGTAGCGTAAAGGTAAAGAGTGA	1875		
Db	1997	TTCTCTTT	TGATATGCTTCCAAAGCCGTACGTGTAGCGTAAAGGTAAAGAGTGA	2056		
Qy	1876	GACAGTGTGA	AGTATTAAGAAATTAAGACCTCATATTAATTAATGAATTTCTTGA	1935		
Db	2057	GATAGGTGC	AAATATCTAATTAATTAAGACCTCATATTAATTAATGAATTTCTTGA	2116		
Qy	1936	AATAATA	CAAAATTAATACAACTGGAAGAAATTTGCTCTTGGCATATGTAATTC	1995		
Db	2117	AATAATA	CAAAATTAATACAACTGGAAGAAATTTGCTCTCTGCGATATGTAATTC	2176		
Qy	1996	AATAGATTTT	CACTAATTTGTCAGATTTTCAAAAAAGATCCTGTACAAAAATTA	2055		
Db	2177	AATAGATTTT	CACTAATTTGTCAGATTTTCAAAAAAGATCCTGTACAAAAATTA	2236		
Qy	2056	AAGCATCGTGA	CCGTTCTATTAAGAAAGCCCTTGGAAGATGCTCTAAAGTAGACCTT	2115		
Db	2237	AAACAATTTGA	AGCCATTTTATGAAGAAAGACCTTTGGAAGACGCTCTCAAGTGAAGCTT	2296		
Qy	2116	CAAGGAAAT	TGACTGATTAATCTAATATGTGAACCTGCATTGAAAGATTAATTTGA	2175		
Db	2297	CGAGGAAAG	CTGAGATGAACAAATGTACTGACGATTTGAAGGATATTTGTA	2356		
Qy	2176	GGGGGAA	CGTAAATATATGCGTCAATCAGTGCTTATGCAAGCTAACCAAGATTAAC	2235		
Db	2357	GGGGGAAAT	TGTAATATATGCGATCACTGAAAGTTTATGCAAGATTAACCAAGATTAAC	2416		

FEATURES  
Source Location/Qualifiers  
1.3081  
/organism="unknown"

ORIGIN  
/mol\_type="genomic DNA"

Query Match 52.9%; Score 1624.8; DB 6; Length 3081;

Best Local Similarity 72.8%; Pred. No. 1.2e-268;

Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

```
1 ATGGCGCGGCGGTGACGCGCAG-----GCAGCAGGACACAGAAATGACATTGAT 51
1 ATGGCGCGGCGGTGACGCGCAGGCTGCAAAAGCATCAGGGCGTAGTGTAATGATGT 60
52 GAGCAACATGTTTACCTTAATCTTAAGAGATGCACTAGTACGACGAGATGCAA 111
61 GAAAGAAATCTTTGGCTTTAATGCA---GGAAAAAATATATGATATGAAATGCAAA 117
112 AAAAACTPAAAAAATATTTGCCAAGATTGACTGAGCAAAAATAATATGAAAGATA 171
118 AAGAAATTGAAAAAATTTGTAAGACATTACGATGCGAATGTAACAGAAAAAGTT 177
172 CACAGAAAACTTAAAGTTTTCGAAAGATGCAAAAGCAGATACAAATGCAAAAGACTG 231
178 CACAAAAAATTAAAGAGTTTGTGAAAAATAAAAAGCAGATTCAAAATGCAAAAGACTG 237
232 AAAGCCAAATTTGAGAAAAATGTACTACATCAAGCAAAAGAACTTAAAGAACATTA 291
238 AAAGAAAAAATCACTCAAAATAATGTCTGCAATCAAAAGAAACCTTACAGAACATCAA 297
292 AAAAAAATTCAGATTATACGAGTAAAGATTGCAAAAGAAATGAACAATGCTATT 351
298 AAAAAAATTCAGATTATACGAGTAAAGATTGCAAAAGAAATGAACAATGCTATT 357
352 TTGAGGAGATGATGTTTCAAAAGAACTTAAAGATGATTGCAATCTTTGAAATTAAGTC 411
358 TTGAGGAGATGATGTTTCAAAAGAACTTAAAGATGATTGCAATCTTTGAAATTAAGTC 417
412 TATCAAAAAGAAAGTGATTAAGTTGCGGAAAGATTTCTTTAAGACACTTGATGCAAT 471
418 TATCAAAAAGAAAGTGATTAAGTTGCGGAAAGATTTCTTTAAGACACTTGATGCAAT 477
422 CTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
478 CTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537
532 GAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591
538 GAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 597
592 ATTGAAAAAGATTAAGATGCGGTACTCTTAAACAGATGTTTCAAGCAGCACTAG---- 646
598 TTGAGAAAAACAAAAAATGTAATGCTCTTAAACAGATGTTTCAAGCAGCACTAG---- 657
647 -GAAATTTTAAAAAAGAAACATGCTGTAATGATGATGATGATGATGATGATGATGATGAT 705
658 AAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
706 AATTCGCGAG-----ACGACGATTAAT 729
718 AATTGGAAGATATATCAAAATGTAATAATCATCGAAGATGTTATGAAATTTGGCA 777
730 AATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 789
738 GTGTGTGATACATATGCGAGTGAATGTAAGAAATATAGTTATTTATACATCCGGGA 837
790 CCGGATTTGATCCAACTAGCGCAAGGCTACATAGCAGAGATATAGGGCTGGAAGAG 849
838 TCCGATTTCAATCCAACTAGCGCAAGGCTACATAGCAGAGATATAGGGCTGGAAGAG 897
850 TTTTATTAAGAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 909
898 CTTTATTAAGAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957
```

```
910 ACACCTTTGTTGGCATTTGATTCAGATTTCTAGTCTTAAAAAAGACGACAAAGAG 969
958 ACTGCTCTACTGGGCGCTTTGATTCAAATCTCTAGATCTTAAGATCAAGTGGTAAAG-- 1015
970 AATGCGAAGAGCCCTTCAAAAAAGCTGCAAAATCCTCATGAAACATGAGCTTTAGAA 1029
1016 -AATGCGAAGAGCTTTCAAAAAAGCTGCAAAATCCTCATGAAACATGAGCTTTAGAA 1074
1030 AGTTATGTAAGAAAAATGTTTAAATGATGAAAGAAAAATGTAAGAAATTCGCA 1089
1075 GATTTTGTATCAAAATGATGCTGTCGCAAAATGAAATGAAATGAAATGAAATGAAAT 1134
1090 AATGATATTAACAACATCTGCAAAATTTTCACTTCAAAAGTCATATATATCGCTTTT 1149
1135 AAGAGTTAGCAACAGATCTTAAATTTCTTTTAAAAAATTAAGAAATTAACACCTCT 1194
1150 GATCCAAACAAAAATTAATGAATTTGATGAGGAAAGGTTGCAACATTTCTTAC 1209
1195 -----GATCCGAGAAAGTCAATTCATGATGATGATGATGATGATGATGATGAT 1242
1210 AACGAAGATGTCGAAATTTGAGTCTTATTTTCTATTTTGAATAAAATGTCAGAT 1269
1243 GACATGATCTGCAACAGTTAGATGATGATGATGATGATGATGATGATGATGATGAT 1299
1270 GGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1329
1300 CTTACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1359
1330 GCAATTAAGTCTGCAAGAAATATGCGAGAGATTTTCACTGTTCAACAAAGCTG 1389
1360 GCTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1419
1390 CTTGAAAGTTTCAACAAAGATTTGATGATGATGATGATGATGATGATGATGATGAT 1449
1420 TTTAAGAAATCTATGAAATAATGATGATGATGATGATGATGATGATGATGATGAT 1470
1450 GGAAGTTCTCAACGATGATTTATTTTCTGATGATGATGATGATGATGATGATGAT 1509
1471 -----ACAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1521
1510 TTGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1569
1522 ATACTTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1581
1570 CGAGATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAATGTCGAAGATTTAG 1629
1582 CGAGATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAATGTCGAAGATTTAG 1641
1630 GAGATTTCAAAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1689
1642 AAGATTTCAAAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1701
1690 GGAATCAAGAAATTTTAAAGCAGGTTTATTTGATGATGATGATGATGATGATGATGAT 1749
1702 GGAATCAAGAAATTTTAAAGCAGGTTTATTTGATGATGATGATGATGATGATGATGAT 1761
1750 CAAGAAATTTGTAATAATTAATTAATAATTAATAATTAATAATTAATAATTAATAAT 1809
1762 CATGAAATCTGTAAGAAATTTTAAAGAAATGATCATTAATGATGATGATGATGATGAT 1821
1810 GACGTTTCTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1869
1822 GATGTTTCTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1881
1870 GTGAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
1882 GTGAAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1941
1930 CTTGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1989
1942 CTTGAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2001
```

1990 TATTGCATAGATTTTTCACCTTAATGTCAGGCTCTTACGAAAGAAATAGT---TGTAACA 2046  
 2002 TACTGCATTAATTTTTCACCAATGTCCTGATCTTAAGAAAGAAAACCTTCTGTGCA 2061  
 2047 AAAATCAAGAACATCGTACCGCTTCTAATAAAGAAAGCCCTTGAAAGATGCTCTCAA 2106  
 2062 AATCTTAAAAAATATGCGAACCAATCTTACAAAAGAAAGGTTTAAAGATGCTCTTAAA 2121  
 2107 GTAGAGCTTCAAGGAAAATGACATGATTAATCTAAATGTAACCGCATGAAAAGATAT 2166  
 2122 GTAAAGCTTCAAGGAAATTAATGATTAATGAATTAATGATGCAATTAAGAAAGATAT 2181  
 2167 TGTAACAGTACGCGGAAACGTAAATATGCTCAATCACTGCTTATATGCAAGCTAACAC 2226  
 2182 TGCAATATTTGAAAAAATGTAAGTATTCATCAATCAACATTTTATGTAAGATTAATCC 2241  
 2227 AAGATTAATCTTGAAAGAGTATGAGATCTTGAAGAAAGAACTCTGTGAGAAATTAAGT 2286  
 2242 GAAAGTAAAACTAAAAAGACCGATATGATGAAAGAAAGCTTTGCTAAAAATTAAGTG 2301  
 2287 AAAAAGTGAAGAAACATGCAAAAGCATTAACCAACAGAAATTAAGCAACCGGACGCTGAT 2346  
 2302 GAAAGGTGAAGAAACGCAATGTAATTAATGTTACAGCAAGAAATGAGCATGAGAGAAAAAGC 2361  
 2347 CTAATAAAAGATTTATAGACATATGAGCACTTAAGAAACGTCGAGAGAAAGCAATGAC 2406  
 2362 CTAATAAGATGATTTTGAAGACATTTGAAAAAATCTTAAAAAAGCGAGAGAAAAACAATGAT 2421  
 2407 AAGTCCAGCTTTGTTTGTCTCACTCATTAAGAAAAAGAAAGTAAATGATCAAAAAGTAAAT 2466  
 2422 AATTCATCTTTGTTTATCATCTGCTTAAGAAAGATGAAATTAATCATCAAAAATTAAGT 2481  
 2467 AGCAAAAACAAAGATTAAGAAATGCGCTTCAACCGAATCTCAAGATTAACCAAAAACATGTC 2526  
 2482 AGCAAAAGCAAGATTAAGAAATGCGCTTCAACCGAATCTCAAGATTAACCAAAAACATGTC 2541  
 2527 AAAATTAACAGAGAGAGAGATTAAGAGATGATCCGTAACAGAAATTAAGAGATTAAGAGATTA 2586  
 2542 AAAATTAACAGAGAGAGAGATTAAGAGATGATCCGTAACAGAAATTAAGAGATTAAGAGATTA 2601  
 2587 GATTGGCAGCAGAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2646  
 2602 GATTGGCAGCAGAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2661  
 2647 GAATTCAGATTCAGAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2706  
 2662 GAATTCAGATTCAGAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2721  
 2707 AATAAGGCTTGTGCAATCTGAAGCCTCTGAGAGTGAAGCCGCAAGAAAGTGAAGAA 2766  
 2722 CAAAGAGTATGTTGAAATTAAGAACTGAAAGTGAAGCCGCAAGAAAGTGAAGAA 2781  
 2767 AGTACAAAGCAATCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2826  
 2782 AGCAAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2841  
 2827 AAATTCCTTACGACAAACGACATGAGGTTTACAGACATGACACACACACACACACACACACACAC 2886  
 2842 AAATTCCTTACGACAAACGACATGAGGTTTACAGACATGACACACACACACACACACACACACAC 2901  
 2887 ACTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2946  
 2902 ACCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2961  
 2947 ACCAAGTACGACAG 3006  
 2962 ACCAAGTACGACAG 3015  
 3007 AGATG 3066  
 3016 AAGATG 3075  
 3067 ATTAG 3072

Db 3076 ATTAG 3081  
 RESULT 11  
 AF033212  
 LOCUS  
 DEFINITION  
 PNEUMOCOCCUS CARINII F. SP. HOMINIS CLONE HUMSG32 MAJOR SURFACE  
 GLYCOPROTEIN (MSG) GENE, PARTIAL CDS.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 ORGANISM  
 PNEUMOCOCCUS JIROVECI  
 PNEUMOCOCCUS JIROVECI  
 Bacteria; Firmicutes; Bacillales; Lactobacillaceae; Lactobacillus  
 PNEUMOCOCCUS JIROVECI  
 1 (bases 1 to 3080)  
 Garbe, T.R. and Stirling, J.R.  
 Molecular characterization of clustered variants of genes encoding  
 major surface antigens of human *Pneumococcus carinii*  
 Infect. Immun. 62 (8), 3092-3101 (1994)  
 94314421  
 751806  
 2 (bases 1 to 3080)  
 Mei, Q., Turner, R.E., Sorial, V., Klivington, D., Angus, C.W. and  
 Kovacs, J.A.  
 Characterization of major surface glycoprotein genes of human  
*Pneumococcus carinii* and high-level expression of a conserved  
 region  
 Infect. Immun. 66 (9), 4268-4273 (1998)  
 98380374  
 9712777  
 3 (bases 1 to 3080)  
 Mei, Q., Turner, R., Sorial, V., Klivington, D., Angus, C.W. and  
 Kovacs, J.A.  
 Direct Submision  
 Submitted (07-NOV-1997) Critical Care Medicine Dept., National  
 Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD  
 20892-1662, USA  
 Location/Qualifiers  
 1..3080  
 /organism="Pneumococcus jirovecii"  
 /mol\_type="genomic DNA"  
 /specific\_host="Homo sapiens"  
 /sub\_species="hominis"  
 /db\_xref="taxon:42068"  
 /clone="HUMSG32"  
 /note="derived from HIV-infected human with P. carinii  
 pneumonia"  
 <1..3029  
 /gene="MSG"  
 <1..3029  
 /gene="MSG"  
 /gene="MSG"  
 /note="surface antigen"  
 /codon\_start=3  
 /product="major surface glycoprotein"  
 /protein\_id="AAC34975.1"  
 /db\_xref="GI:3560521"  
 /translation="AAVAKRQAAKASGASVYDGEIILALIAKRYNDNECKELEK  
 CHTLTDALPEKVKHKLKEFCENKSKDCKEIKELCTOKCTAIAKLTAEAKKNS  
 DLTDDEKENEQCLPLEGACPAELKDDCTLEBSDFYLSQAPLDKCNLKAQYAGLE  
 NRTTBEGKLKEVCIEISDESDDELTKLCLYKMTCTFVLRKDKCNLAKODVNALEK  
 KDELRGKCLPLBRVCYPRNCEPDISKCNSEDECVYLVCOTLAKEENK1IYH  
 PGSDPFPKSKPTAEDIGIBELRYKKAESGVHIGKRPVADATLALLIIONDPKO  
 VGRCEBVLNDCKELKSHILGDFQNVAGQNEIKCKELEKELANSTKILYEKLE  
 NKHLSGSEVAPYWKLTPLSDNDCTLEBSDFYLSQAPLDKCNLKAQYAGLE  
 AQAIEALQKMYGILFYSGKWEFKLLEKIMECESEKTTSDLEFLICIDPLKAVRIL  
 AADTQARAVALFKOLDKDPDPTDKCEIKCKEALGDSNOIKMFCITLKQCCDL  
 GTTEILKQVLLDEHKOTLRHENTCTYKLRCKHWSRRGDRSPFVGVONATCKLIV  
 DTVKDRCEVBEKMKQASDINSLKDKITKESANIPSPHNPICDRPLPNCPODKKK  
 TPCQNLKRYCEPFRKRVLEDAKVEIQSLSNKCESALERTCTILKVSBSINS  
 LCKDNTESKTKTDNEVRKQLCLKLVVEVQCCMKLPALDEHEKDKLDFETEEKL  
 KQAEKTMKNSNLVLSFVKDEBNNTSKNSKDKDKNTVSNGLQDTTEHMKILRRGVKV



SVTSEAKAFDVAEFGRYLDKERCKLSDCKVKBDCXDLBSGVCKIGVCSKIK  
 PLKVPHEVTESTITTTTTTTTTTVDTPATECKSLQTDWTWITQSTHSTSTST  
 ISKITLSTRRCKPKTKTGDADVKSEBGLKMSG"  
 misc\_feature 3029  
 /gene="MSG"  
 /note="putative PCR artifact leads to premature stop  
 codon"

## ORIGIN

Query Match 52.9%; Score 1623.8; DB 8; Length 3080;  
 Best Local Similarity 72.8%; Pred. No. 1.8e-268;  
 Matches 2274; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

2 TGGCCGCGCGGTCAAGCGGCAG-----GCAGCAGGACACAGAAATGCAATTG 52  
 1 TGGCCGCGCGGTCAAGCGGCAGGCTGCAAAAGCATCAGGGCTGTGATATGATGCTG 60  
 53 AGGACATGTTTACTTTAATTCTTAAGAGATGACTAGTAGAGAGAGATGCAAAA 112  
 61 AAGAAATTCCTTTGCTTTAATTGCA--GGAAAAAATATATATATATGATGCAAAA 117  
 113 AAAAACTAAAAAATATGCAAGATGATGCAAGCAAACTAATATAGAACAGTAC 172  
 118 AAGATTGAAAAAATTTGTAGACATTAAAGGATGCAAAATTAACCAAGAAAAGTTC 177  
 173 ACAGAAAACTTAAAGGTTTTCGCAAGATGAAAAAGCATACAAAATGCAAAAGTGA 232  
 178 ACAAAAAACTTAAAGAGTTTGTGAAAAATAAAAAGCATTCAAAATGCAAAAGATGA 237  
 223 AAGCAATATGAGAAAAATGATCTACATCAAAAGAAAACTTAAAGAGCAATTA 292  
 228 AAGAAAACTCTCAAAAATGATCTGCAATCAAGAAAACTTACAGAGCAATCAAAA 297  
 293 AAAAAATTCAGATTATTAAGATGAAAGATTCGAAAGAGATGAACAATGCTATTTT 352  
 298 AAAAAATTCAGATTATTAAGATGAAAGATTCGAAAGAGATGAACAATGCTATTTT 357  
 353 TGGAGGAGTATGTTCAAAAAGAACTTAAAGATGCAATCTTTGAGAAATTAAGTCT 412  
 358 TGGAGGAGCATGTCAGCGGAACTTAAAGATGCAATCTTTGAGAAATTAAGTCT 417  
 413 ATCAAAAAGAAAGTAAAGTTGCGGAAAGTCTTTTAAGACACTGTGTAGCATC 472  
 418 ATCAAAAAGAGCTAAAGTTGCGGAAAGCTCTTTTAAGACAGTTCTGTAGAGTTC 477  
 473 TTAATGATCACTATGCTGAAAAAATTTAAAGATTTGCCCTGTCAATGCGGAGG 532  
 478 TAAATCAATGAATCTCATGTGAAGAAAGCTCAAGAGGTTTGAATGAGTCAAG 537  
 533 AAAGTATGATTTAACTTAAGCTTGTCTGACCAAGAAAGACATGTAAGAAATTTTAA 592  
 538 AAAGTATGATTTAACTTAAGCTTGTCTTATCAAAAATGACGTGCAAAACATTTGTAT 597  
 593 TTGAAAAAGATTAAGAGTGCCTTCTTAAACAGATGTTTACAGACAGCTAG----- 646  
 598 TGAAGAAACAAAAAATGTAATGCTTTAAACAGATGTTAAACGAGACCTTGAGAAA 657  
 647 GAAGTTTAAAAAAGAAATGCTTGAATTAATGCAACATGCTATTTTTCATTGGA 706  
 658 AAGATGAGTTTACGAGGAAATGTTTACCACTGCTGAACGATGCTATTTTATAGGGA 717  
 707 ATTGGGAG-----ACGACGATATTA 730  
 718 ATTGGAAGATATATCAAAATGTAATTAATCATCCGAAGACTGTTATGATTTTGGCAG 777  
 721 AATGATTTGAATTTGGAGGAAAAATGCAAGAACAAATTTGCTTATATGCAACAGAGC 790  
 778 TGTGTGATTCATTTGGAGGAAATGGAAGAAATTAAGATTAATTAATCAATCCGGAGT 837  
 791 CCGATTTTGAATTCATGAGGCAAGGCTTACATATGACAGAGATTAAGGCTGAGAGT 850  
 838 CCGATTTTCAATCACTAAGTCAAGGCTTACTGTAGCAGAGACATATGAGACTGAGAGC 897

851 TTTATTAAGAGTGAAGAGATGAGTATTTTATTGAAAAAGATCATCTAGAGATGGCA 910  
 898 TTTATTAAGAGGCGCAGAAAGAGTGTATATTTGAAAGCCCTCTGTAGAGATGCA 957  
 911 CAGCTTTGTTGGCATTTGTTGATTCAGATTTCTAGTCTTAAAAAAGAGACAGCAAGAGA 970  
 958 CTGCTTACTAGCCCTTTTGAATTCAAAATCTAGATCTTAAAGATCAAGTGGGTAAAG--- 1014  
 971 AATCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAAATGAGGCTTTAGAA 1030  
 1015 AATCGAAGAGGTTCTTAAAGATTAATGTAAGATTTAAAGATCAAGAAATTTTGGAG 1074  
 1031 GTTATGTAAGAAAAATGCTTAAAGTAAATGATGTAAGCAAGAAAAATGCAAAATTCGAAA 1090  
 1075 ATTTTGTAAACAAAATGATGCTGTCAAAATGAAATTTGAAAGATGTAAGATGAGA 1134  
 1091 ATGATATTAACAAAATCTGCAAAATTTTCACTTCAAAAGTCAATTAATGCTCTTTTG 1150  
 1135 AGGATTAAGCAACATGCTTAAATTTCTTTTGAAGAAAAATTAAGATTAACACTCTCT- 1193  
 1151 ATCCAAAGAAAGAAATTAATGAAATTTGATGGAAGGAGGTTGCCAATTTCTTACGA 1210  
 1194 -----GATCCGAGAGATCAATTCATGTTAAAGTTGACGACATTTCTTAGTG 1242  
 1211 ACAGAGATTTGCGGAAATGGAAGTCTATTTGTTCTATTTGAAAAAATATGTCAGATG 1270  
 1243 ACAATGACTGCAAGGTTAGAGTCACTGTTTTTATTT--AAAAATCAAGACCTC 1299  
 1271 GAGAAAAATGCAATGAAAAATTAATTAAGACCAATGTTCAAAAAGAGACTTATGATCAGG 1330  
 1300 TTGACAAAAGATGTAATTAATCTGAAGCAGACATTTATTAAGAGGCTTGAACAGACAG 1359  
 1331 CAAATTAAGTCTGCAAGAAAAATTAAGAGAAATGTTACATGTTCAAAACAAAAGCTGGC 1390  
 1360 CTAATGAAGCAATGCAAGAAAAAGATGACGACTGTTCTATGTTTCAAGCAAGATGCT 1419  
 1391 TTGAAAAATTTCAACAAAGATTAATTAAGATTAATGTAAGAAATCTGAAAAAAGAAAAA 1450  
 1420 TTAAGAAATTAATGAAAAAATTAATGAAAGATTTGCGAACTTAAACA----- 1469  
 1451 GAAGTTTCTCAAGATGAAATTAATTAATCTGTGTGACAGCCAGCAAAAGACGCCGT 1510  
 1470 -----ACAAGGATGAGTGTGTTTGTCTATGATTAAGTCCACTTAAACAGTCAAG 1521  
 1511 TGCTTACATGATCTTGAATGAAAACTATCTTTTATGACAACTGATCAAAAGC 1570  
 1522 TACTTGACATGATATCCAAAGACAGACGCTTTTGGGAAACAAATGATCAAAAGC 1581  
 1571 GAGATTTCCGACAGATTAATAATGCAAGAAATTTGGGAGAAAGTCCAAATTAAGAG 1630  
 1582 GAGACTTCCAAAGACAAAGATTTGCAAGAAATTAAGAAAGATGTAAGCTTTAGGGA 1641  
 1631 AGGATTCAAAAGAAATTAATGAGCATGTCATACCTGAGACAGCAATGCAATGCTTGG 1690  
 1642 AGGATTCAAAATCAATTAAGTGGCAGTCTATGCTTAAACACAGTGTGATGCTTGG 1701  
 1691 GAGCTACGAAATTTTAAAGCAGGTTTATGATGAACACAAAGATTAATTTGAAAGAC 1750  
 1702 GAGCTACGAAATCTTGAACAGGTTTATCAATGAGAACAAAGATCTTTAAGAACTC 1761  
 1751 AAGAAAGTGTGTAAATTAATCTTAAAGAAAGTGAATTAATGCTTGAAGAGAGATG 1810  
 1762 ATGAAACCTGTAAGAAATTAATTAAGAAATTAATTAATGCTTGAAGAGGCTGATG 1821  
 1811 ACCGTTTCTTTTGTATGCTCTTCCAAAGCCTAGCTGTGAGCTGTTAAAGAGC 1870  
 1822 ATGCTTTCTTTTGTATGCTTTTAAACAAAGCCTAGCTGTGATGATGATGATG 1881  
 1871 TGAAGACAGGTTGAGATTAATTAAGAAATTAATTAAGCTTCAATTAATTAATTTG 1930  
 1882 TGAAGACAGGTTGAGATTAATTAAGAAATTAATTAAGCCTGAGATTAATTAATTTCT 1941  
 1931 TTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1990

D	b	1942	TTAAAAATTAACAATTAATAACAGATACAGACGAAATATTTGTCCTCATGCAACCAT	2001
Q	y	1991	ATTGCAATAGATTTTCACTTAATTTGCCAGTCTTACGAAAGAAATAGT--TGACAA	2047
D	b	2002	ACTGCGATAGATTTTATCCCAATTTGTCGTATCTTAAGAAAGAAAACTTTCTGTCAA	2061
Q	y	2048	AAATCAAGAACATCGTAGAGCCGTTCTATAAAGAAAGCCCTGGAAAGTGTCTCAAG	2107
D	b	2062	ATCTTAATAAATAATTTGGAAACCAATTTCAAAAGAAAGGTTTGAAGATGCTCTTAAAG	2121
Q	y	2108	TAGAGCTTCAAGAAAATTTGACTGATAATCTAATGTGAACCTGCAATTTGAAAAGATTT	2167
D	b	2122	TAGAGCTTCAAGGGAATTTAAGTAATGAATAATGTGAATCTTGCAATTAGAAAAGATTT	2181
Q	y	2168	GTACAGTAGCGGGAAACGTAAATATGCGTCATCAAGTGGCTTATGCAAGCTAACCA	2227
D	b	2182	GCACATATTTGAATAATGTAGTATTCATCAATCAACAGTTTATGTAAAGATTAATCCG	2241
Q	y	2228	AGGATTACTCTGGAAAGATGATGAGATGCTAGAAAGAACTCTGTGAGAAATTAAGTA	2287
D	b	2242	AAAGTAAAACTAAATAAGACCCGATATGAAAGTTAGAAAGAACCTTTGCTAAATAATTAAGTG	2301
Q	y	2288	AAGAAGTGAAGAACAGTGCAAACATTAACAAGAAATTAGCAACCCGACACTATC	2347
D	b	2302	AAGAGGTGAACACCAATGTAAATGTTTACAGACGAATTTGAGACATGAGGAAAAAACC	2361
Q	y	2348	TAAAAAAGAATTAATAGACATATGAGCACTTAAGAAAACGTGACAGAGAAACATGAA	2407
D	b	2362	TAAAGATGATTTTGAACAATTTGAAAACCTTAATAAACAAGCAGAGAAAACATGAAATA	2421
Q	y	2408	AGTCAGTCTTGTTTTGTCACTCATTAAGAAAAAGAAAGTATGTATCAAAAAGTAAATA	2467
D	b	2422	AATCAATCTTGTTTTATTCATTCGTTAAAGAAAGTGAATAATATATCATCGAAAATATGTA	2481
Q	y	2468	GCAAAAAACAAGATTAAGAAATGCCCTTCAACCGACCTTCAAGATATCCACAAACATGTGA	2527
D	b	2482	GCAAAAGCAAGGATTAAGATACCGTTTCAACCGACTTCAAGATATCCACAGAAATATGTA	2541
Q	y	2528	AAATACTAACGAGGGAGATTAAAGATGTATCCGTAAACGAATTTGAAGCTTAAACATTTG	2587
D	b	2542	AAATACTAACGAGAGAGATTAAAGATGTATCCGTAAACGAATCTGAAGCTTAAGCATTTG	2601
Q	y	2588	ATTTGGACAGCAAGATATTTGGAAAGATATGTAGATTTGAAGAAAAGATATATTAATTTGG	2647
D	b	2602	ATTTGGTAGCAAGATATTTGGAGATATCTPACCTTGAAGAAAAGATGTATTAATTTGG	2661
Q	y	2648	AATACAGATTGCAGAAATTAAGAGGATTTGCCAAAGACTTAGAAGAAAGTATGCAAAAAGTTA	2707
D	b	2662	AATCAGATTTGCAGAGTTAAGAGGATTTGCCAAGATTTAGAAAGAGTATGTGGAAGATATC	2721
Q	y	2708	ATAAGGCTTGTGCGAATCTGAAAGCCTGCGAGGTGAAGCCGACAGAAACAGTGCAGAAA	2767
D	b	2722	AAGGATATATTTGCAAAATTAABACCACTGAAAGTGAAGCCGACAGAAACAGTGCAGAAA	2781
Q	y	2768	GTACAAAGCAACTTACAAACAACAACAACCGTTGCCGATCCGAAGGCAACGGAATGCA	2827
D	b	2782	GCACAAAGCAAGCAACGACGACAAACAACGATCGTTATCGATCCGAAGGCAACAGATGCA	2841
Q	y	2828	AATTCCTTAACAGACAACAACAACATGCGTTACACAGACATGCAACAACAAGCACGTCTA	2887
D	b	2842	AATCTTTTACAGACAACAAGATTCATGTGATTAACAAGACTTTCACACATATCCAGCACGTCTA	2901
Q	y	2888	CTATCAACATCTCAACATCAATCAAAATAATCAATTTGACATCAACGAGGGAATGCAAAACAA	2947
D	b	2902	CCATCAACATCTTACACATCAACATCAAAATAATCAACTCACTCAACAAGGCGTTGCAAAACAA	2961
Q	y	2948	CCAAGTGTACAGACAGAGAGGAAGATGTGACAGAGACGTGAAAACCGAGTAGAGGCGCTGA	3007
D	b	2962	CCAAGTGTACAGACAGG-----GGATGATGCAAGAGACGTGAACCGAGTGAAGGAAATTGA	3015
Q	y	3008	GGATGATGGGTGGATATGTGATGAGGGGGGTGATATGTAGCAATGTTATTTCTTCATGCA	3067

Db	3016	AGATGACTGGGTAAACGTATGAGGGGGGTGATAGTACGAAATGGTATATTTCGTTATCA	3075
Cy	3068	TTTAG 3072 	
Db	3076	TTTAG 3080	
RESULT 12			
LOCUS	AF038556	12792 bp	DNA linear PLN 10-SEP-1998
DEFINITION	Pneumocystis carinii f. sp. hominis variant regions of major surface glycoproteins (meg1, meg3, meg4) genes, partial cds.		
ACCESSION	AF038556		
VERSION	AF038556.1	GI:3560524	
KEYWORDS			
SOURCE			
ORGANISM	Pneumocystis jirovecii Pneumocystis jirovecii Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;		
REFERENCE	1 (bases 1 to 12792) Met, O., Turner, R.E., Social, V., K Livingston, D., Angus, C.W. and Kovacs, J.A.		
AUTHORS	Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region		
TITLE	region		
JOURNAL	Infect. Immun. 66 (9), 4268-4273 (1998)		
MEDLINE	98380374		
PUBMED	9712777		
REFERENCE	2 (bases 1 to 12792) Met, O., Turner, R., Social, V., K Livingston, D., Angus, C.W. and Kovacs, J.A.		
AUTHORS	Direct Submission Submitted (12-DEC-1997) CCMD, NIH, Building 10, Room 7D43, MSC 1562, Bethesda, MD 20892-1662, USA		
TITLE	location/Qualifiers		
JOURNAL	1. 12792		
MEDLINE	/organism="Pneumocystis jirovecii"		
PUBMED	/mol_type="genomic DNA"		
REFERENCE	/specific_host="Homo sapiens"		
AUTHORS	/sub_species="hominis"		
TITLE	/db_xref="taxon:42068"		
JOURNAL	<668. 3712		
MEDLINE	/gene="meg1"		
PUBMED	/note="multicopy gene family is expressed by gene rearrangement"		
REFERENCE	<668. 3712		
AUTHORS	/gene="meg1"		
TITLE	/note="multicopy gene family is expressed by gene rearrangement"		
JOURNAL	/note="glycoprotein A; expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"		
MEDLINE	/codon_start=1		
PUBMED	/product="variant region of major surface glycoprotein"		
REFERENCE	/protein_id="AAC34980.1"		
AUTHORS	/db_xref="GI:3560525"		
TITLE	/translation="VAAVAVKQVTGASGVDBEEVRLALILKEDSKDDKCEBTELEKHC CKEISEANLPPEQVHEKLDKPCDSKKRDKCKELKXNVEKCCDFTLEBELVKEKAS NLKNDCEKNEOCLPLEBASCDLTGKNDLIRKCYONKDKTAKAEVTLILIKGKNPK DKNSCEKLEEVYCOELSDMSDEMLKCFDQNTCDNLVKETQOCSEFKNILTEIKTI KEBOLKCKKPLIYEBCTFYDESGCNDLSKSELEKCOEKNITLTVSYSGPDEBE TTLAEBVDLSGTYRKAABEETVQKPLPADATLVAFLIDPSLTORTNKKCKTIL EDCKLAKHNDIITKGLCEDYTNANDKDKCEBSTDIEETCKFFISKTLIHFFDGN KNQDILKMNISTPLSNKDCITLSEYCLCYEKSCSRBTACKNIRAAQYKGLDTLANE VLQKEMRGMHSGNKITLISGFQKLLLEVCKVKKENGVAPSNELPYLCVQSPKVAL LSHDLRKTIFLQDLIRKRDPIYKECEBLLKKECALRKDSKIEWPCTLSQNCQ LRNAKELKELLNENQDILKNOENCGVILEKKNEMSRNRNERPSLICALONTGRIIM VEDVANDCKIFEKNIKTYGSDIKTKTIBELGTCTPIHPRCHRGPRCKPLEKXCD PEYCKEYKQRLDENALIFEPHGHDKKKNCKNTNDKYCTLMDQYNTKTLKPFONS TDNNETFRDLCEKLVORVEKCOGLSKELKAKNDLEEHQKDYQKQNTKAMEST NLVSTKSTNDKTEKGVKSTPSVQDILVPLVYKNEKVOTEKAKAFDILVALF SLVLETCKTHLKDCEPFEKCKQCKQCEIKELCKLPELVKPEIKVTEINITT TVETVVEAKTYGDKGKSLSTTDVWVKTSTHSTSTSTTSVTSRILVTSRCK PRCTTDEBDAEAGVKSSEGLSRMSGMSVAGVLLAMISMTI"		
misc_feature	4230..7225		

```

gene
/note="encodes a major surface glycoprotein but due to a
frame shift has a premature termination; correction of the
frame shift results in a complete open reading frame for
meg2".
>7735..10743
/feature="marg3"
/note="multicopy gene family is expressed by gene
rearrangement"
>7735..10743
/feature="marg3"
/note="glycoprotein A; expressed from a single putative
leader sequence located upstream of coding region; variant
region is in frame with the leader sequence"
/codon_start=1
/product="variant region of major surface glycoprotein"
/db_xref="gi:3560526"
/translation="VARAVKRRRAAONSVEEYLLILNEYENNDDKKKRLKEYCE
VLKRVTEPKRLEKLDGICDDDKTIRAKCESTTKRAKTSFQRIEDLVAKKSGST
LEDDCKNERKQCLFLEGACPTBMDKCNELRNKCYQKRDVAEKLLVLEGNLND
KNTKCNLKVGOEFENKSEDLNLCLDEKERTGDIVSKYEKCPKLEPIDLVARE
DLLEKCLFLBECYFYGNSCETDQPKCEFAASCOENLVYAPGSHPEPTLKLRL
AEEDLEKLVYEAVKGIHIGRPISIKOEVLNLLLSGDQNEFKDCEVIVKQGN
FKHEHILKDLCSNKTITDNPKEKEIKELNLTIRILVSRIRKRYAPAVEKILQMH
MLHFLGEBECTKLISCPFIKSAQAPLERKCNILKAQYKGLAEVANBLQRLKSG
LQSGNRNRTVLEQLNLKVCCEKTKGSEDLFVLVCPMNIKALTVSDTLRRVAALQEN
LNEGRDPTPEKDEKLEKCEVILQKDEIRELFWSCYTLQKCNLKLSEHLEIDLVKE
KGYLJENSCKEBAKRCCEPAKRENNKFPASACDELVEKLTITRYVESKNTLHKM
ETNVVISEIAKEKEKICEPAPYCKTQKQNECNKNGKQCKKLNKCKSELEBA
LENVVEIKGSLSNVEGCNNTNLYCTOLKKENGELFTLCKSKNTKSPDIRRELC
EKLIRIKKCKSKDLDEBYEKLVELEKBEKTKIKSEAKAMEDANLILSRKQEDN
NNNSVANDSDPTREKSGITGFVLVRNNAKVHTEKELAFDLVAAFDLYLEKEI
CNHLLKNGCFKKECDCEDPCKKIQGISLEPLKVRHEILVTNLIITVTTTTTTLK
DAKATDCHSLQTTDTWTVKITSTSTSTSTSVTSVTSIRCTKRCPTKGTSEBDA
GDVPSSEGLRMSGVSWVRGVLAMTISFMI"
>11268..>12792
/feature="meg4"
/note="multicopy gene family is expressed by gene
rearrangement"
>11268..>12792
/feature="meg4"
/note="glycoprotein A; expressed from a single putative
leader sequence located upstream of coding region; variant
region is in frame with the leader sequence"
/codon_start=1
/product="variant region of major surface glycoprotein"
/db_xref="gi:3560527"
/translation="VARAVKRRRAAOTNSIDEHVALLIKENGLSFOECKKLKYCC
QELTEAKNLNEQVHAKLKGFEEDGKATKELKELVANIERTCTIKGLIKELAIRKNSD
LTDLEAKNLNEQVHAKLKGFEEDGKATKELKELVANIERTCTIKGLIKELAIRKNSD
SVIEKLLKEICPVMGRSEDELITNLCSNOKETCNILIEKDKKCGTLKTVSAALSF
KKEICLLELCEQCYFYGNCGDDDIICIEILGKQEOBNIAVMPGPDPTREATAI
EDILGEEPKYKVEEDGVPIGKINHLRDAITALLIIDSLSKKDDXKCEALQSKCK
NPHEHLESELCKKNGSLNDSTKCEKLEQNDINKTCITPFSKYTNRLRGLPTGNNBI
VNGHGLPFLPSNEDCAKLESYCFYFEKKCPDGEAKCNIRATYKGLDRAKAVIQE
NMRMLHGSNKSMLKEFOELVVKCEFLKKNKGSFSDDELFTLCVQPAKARLLTHD"

ORIGIN
Query Match 48.8%; Score 1498.4; DB 8; Length 12792;
Best Local Similarity 99.0%; Pred. No. 4.5e-247;
Matches 1508; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

2 TGGCGCGGGCGGTCAAGCGCAGAGGAGCAACAAGATAGATTGATGAGGAACATG 61
|||||
Db TGGCGCGGGCGGTCAAGCGGAGGCGCAGAGGAGCAACAAGATAGATTGATGAGGAACATG 11329
|||||

62 TTTTACGTTTATTTCTAAAGAAATGAGACTAATGTAGACAGAAATGCAAAAAAATCTAA 121
|||||
Db TTTTACGTTTATTTCTAAAGAAATGAGACTAATGTAGACAGAAATGCAAAAAAATCTAA 11389
|||||

122 AAAAAATATTCGCAAGATTTGACTGAAGCAAACTAAATATAGAACAMGTACACAGAAAC 181
|||||

```

Db	11389	AAAAATATATCCCAAGAAATTTGACTGGAAGCAAAACTAAATATAGAAACAAGTACACAAAAAAC	11446
Qy	182	TTAAAGCTTTTTCGGAAGATGAAAAAGCAGATACAAAAATGCAAAAGAACTGAAGCCAAATA	241
Db	11449	TTAAAGCTTTTTCGGAAGATGAAAAAGCAGATACAAAAATGCAAAAGAACTGAAGCCAAATA	11506
Qy	242	TTGGAAGAAAAATGTACTACATCAAAAGAAAACTTAAAGAGCAATTTAAAAAAAATTC	301
Db	11509	TTGGAAGAAAAATGTACTACATCAAAAGAAAACTTAAAGAGCAATTTAAAAAAAATTC	11566
Qy	302	AGATTATTAAGGGATTAAGGATTGCAAAAGAGAAATGACAAACAATGCCATTTTTCGAGGGAG	361
Db	11569	CAGATTTAAGGGATTAAGGATTGCAAAAGAGAAATGACAAACAATGCCATTTTTCGAGGGAG	11622
Qy	362	TATGTTCAAAAAGAACTTAAAGATGATTCGAATACCTTTGAGAAATTAAGTCTATCAAAAGA	421
Db	11629	TATGTTCAAAAAGAACTTAAAGATGATTCGAATACCTTTGAGAAATTAAGTCTATCAAAAGA	11688
Qy	422	AACCTGATTAAGTTTTCGGAGAAAGTTCTTTTAAAGACACTTCGTAGCGATCTTAAATGAT	481
Db	11689	AACCTGATTAAGTTTTCGGAGAAAGTTCTTTTAAAGACACTTCGTAGCGATCTTAAATGAT	11746
Qy	482	CAGTCATATGTAAGAAAAAACTTAAAGAGATTGGCCCTGTCATGGGAGGAGAAAGTATG	541
Db	11749	CAGTCATATGTAAGAAAAAACTTAAAGAGATTGGCCCTGTCATGGGAGGAGAAAGTATG	11806
Qy	542	AGTTAACAAACTTGTGCTGACACGAAAGAGACATGTAAAGATATTTTAAATTTGAATAAG	601
Db	11809	AGTTAACAAACTTGTGCTGACACGAAAGAGACATGTAAAGATATTTTAAATTTGAATAAG	11866
Qy	602	ATTAAGAAATGCGGTACTCTTAAACAGATGTTTCACAGACATCGAAGATTTTAAAAAAG	661
Db	11869	ATTAAGAAATGCGGTACTCTTAAACAGATGTTTCACAGACATCGAAGATTTTAAAAAAG	11928
Qy	662	AAATATGCTTGAATTACTCGAACATGCTATTTTAACTATGGAATTCGGGAGAGAGAG	721
Db	11929	AAATATGCTTGAATTACTCGAACATGCTATTTTAACTATGGAATTCGGGAGAGAGAGAG	11986
Qy	722	ATATTAATTAATATGTAATGGAATTCGGAGAAATGCGCAAGAACAAATATTTGCTATATGCG	781
Db	11989	ATATTAATTAATATGTAATGGAATTCGGAGAAATGCGCAAGAACAAATATTTGCTATATGCG	12046
Qy	782	CACCAAGAACCCGATTTTGATCCAACTTAGGCGCAGAGGCTCAATATAGCAGAGATATAGGCG	841
Db	12049	CACCAAGAACCCGATTTTGATCCAACTTAGGCGCAGAGGCTCAATATAGCAGAGATATAGGCG	12106
Qy	842	TGGAAGAGTTTATTAAGAAAGTGAAGAGATGGAATTTTATTTGAAAAGAAATCATCTAA	901
Db	12109	TGGAAGAGTTTATTAAGAAAGTGAAGAGATGGAATTTTATTTGAAAAGAAATCATCTAA	12166
Qy	902	GAGATGCGACAGCTTTTGTGGCAATGTTGATCCAAAGATTTCTAGCTTTAAAAAAAAGAGCG	961
Db	12169	GAGATGCGACAGCTTTTGTGGCAATGTTGATCCAAAGATTTCTAGCTTTAAAAAAAAGAGCG	12222
Qy	962	ACAAAGAGAAATGCGAAGAAAGCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAACATGAGG	1021
Db	12229	ACAAAGAGAAATGCGAAGAAAGCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAACATGAGG	12288
Qy	1022	CTTTAGAAAGTTTATGTAAGAAAAATGCTTAACTAATGATGGAACGAAAAAATGTGAAG	1081
Db	12289	CTTTAGAAAGTTTATGTAAGAAAAATGCTTAACTAATGATGGAACGAAAAAATGTGAAG	12346
Qy	1082	AATTCGAAAAATGATATTTAAACAAAACCTGCAAAATTTTCACTTCAAAAAGTCACTAATATTC	1141
Db	12349	AATTCGAAAAATGATATTTAAACAAAACCTGCAAAATTTTCACTTCAAAAAGTCACTAATATTC	12406
Qy	1142	GTCCTTTTTCATCCAAACAAAAGAAATATGAAAATGTTGGATGGAAGGGTTCACAACAT	1201
Db	12409	GTCCTTTTTCATCCAAACAAAAGAAATATGAAAATGTTGGATGGAAGGGTTCACAACAT	12466
Qy	1202	TTCTTAGCAACGAAGATTTGCGAAATTGGAGTTCATGTTTCTATTTTGAAAAAAAT	1261



Db 12469 TTCTTACCAAGAGATGTGCGAAATTAGAGTCTTATGTTCTATTTTGAATAAAAT 12528  
Qy 1262 GTCCAGATGAGAAATATGCAATGTAATTAAGACATGTTAATAAAGAGACTTG 1321  
Db 12529 GTCCAGATGAGAAATATGCAATGTAATTAAGACATGTTAATAAAGAGACTTG 12588  
Qy 1322 ATGCAAGGCGCAATTAAGTGTGCAAGAAATATGCAAGATGTTACATGTTCAACA 1381  
Db 12589 ATGCAAGGCGCAATTAAGTGTGCAAGAAATATGCAAGATGTTACATGTTCAACA 12648  
Qy 1382 AAAGCTGCTTGAAGATTTCACACAGATTATGTAAGATGTGAGAACTGAAAAAG 1441  
Db 12649 AAAGCTGCTTGAAGATTTCACACAGATTATGTAAGATGTGAGAACTGAAAAAG 12708  
Qy 1442 AAAACAAAGAGATGTTTCAACAGATTATGTAAGATGTGAGAACTGAAAAAG 1501  
Db 12709 AAAACAAAGAGATGTTTCAACAGATTATGTAAGATGTGAGAACTGAAAAAG 12768  
Qy 1502 CAGCCCGTGTCTTACACATGATC 1525  
Db 12769 CAGCCCGTGTCTTACACATGATC 12792

RESULT 13  
AR438434 3042 bp DNA linear PAT 20-FEB-2004  
LOCUS AR438434  
DEFINITION Sequence 1 from patent US 6664053.  
ACCESSION AR438434  
VERSION AR438434.1 GI:42663289  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 3042)  
AUTHORS Kovacs,J.A., Huang,S., Maser,H., Fischer,S.H., Gill,V.J. and Mei,Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG)  
JOURNAL gene of human Pneumocystis carinii  
Patent: US 6664053-A 1 16-DEC-2003;  
FEATURES  
Source Location/Qualifiers  
1..3042  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 35.8%; Score 1098.6; DB 6; Length 3042;  
Best Local Similarity 63.0%; Pred. No. 1.9e-178;  
Matches 1954; Conservative 0; Mismatches 1054; Indels 93; Gaps 13;

Db 362 AACATGCTC---TGATCTTACAAAGATGCAACGATTTAAGAAACAAATGTTATCAGA 418  
Qy 419 AGAAAGCTGATAAGTTGCGGAGAAAGTTCTTTAAGAGACT---TCGAGAGATCTTA 475  
Db 419 ATAAAGCTGATAAGTTGCGGAGAAAGTTCTTTAAGAGACT---TCGAGAGATCTTA 478  
Qy 476 ATGATCATGATATGTAAGAAATTAAGAGATTTGCGGATGAGAGAGAGAGAGAGAG 535  
Db 479 AAGTAAATTTTATGTAAGAAATTAAGAGATTTGCGGATGAGAGAGAGAGAGAGAG 538  
Qy 536 GTGATGATTAACAACTGTTCTGAAACCAAGAGACATGTAAGATATTTTAATTG 595  
Db 539 GTGACGAATGATGAATATGTTTGAATCAAAAAATAGTGTGATTAATCTGTAAAG 598  
Qy 596 AAAAGATTAAGAGT-----CGTACTCTTAAACAGATGTTTCAGCAGACTAG 646  
Db 599 AAACGCAACAAAGTGAATCTTTCAAAATCTTAAACGGAATTTAAACATTAAG 658  
Qy 647 GAAGTTTAAAGAAACATGCTTGAATTAATTAATGTAAGTTGAGAGAAATGCCAGACAA 706  
Db 659 AAGTGAACATTAAGAAATTAAGTTTATTAATTAAGAGATGATTTTATGATG 718  
Qy 707 ATTGCGGAGACGAGATATTAATTAATGTAAGTTGAGAGAAATGCCAGACAA 766  
Db 719 AAAGTTGTAAGAAAGATTAATTAATTAATGTAAGTTGAGAGAAATGCCAGACAA 778  
Qy 767 ATATGCTTATTAAGTCCAGACCCGATTTTGAATCCACTAGCCGAGGCTTCAATAG 826  
Db 779 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 838  
Qy 827 CAGAGATATTAAGGCTGGAAGATTTTAATTAAGAGTGAAGAGATGAGATTTTATG 886  
Db 839 CAGAGATATTAAGGCTGGAAGATTTTAATTAAGAGTGAAGAGATGAGATTTTATG 898  
Qy 887 GAAAGATCATCTTAAGAGATGCAACAGCTTTGTTGATTTGATCCACTAGCCGAGGCT 946  
Db 899 GGAACCTTTACAGCAGATGATGCTAGCTTTGAGGATTTTGAATCAAGATCCATCTC 958  
Qy 947 TTTAAA---AAAAAGCAGAAAGAAATGCCAGAGAGCCCTTCAAAAAAGCTGCAAA 1003  
Db 959 TTTCAACTCAAGCACTTAAGAAAGAAATGCCAGAGAGCCCTTCAAAAAAGCTGCAAA 1018  
Qy 1004 ATCTCATGAACATGAGGCTTTGAAGATTTATGTAAGAAATGTTTAAGTATGATG 1063  
Db 1019 ATTTAAAGACATGATATTAATTAAGGCTTATGCGAGATTTATATGCAATTAAGATA 1078  
Qy 1064 GAAAGAAATATGGAAGATTTGCAAAATGATATTAACAACTTGCAGAAATTTTCACTT 1123  
Db 1079 AGACAAATTAAGGAGAACTTAAGTACAGATTAATTAAGAAATGATTTTCACTT 1138  
Qy 1124 CAAAGTCACTAATATGCTTTTTCATGATCCAAAGAAATTAAGAAATTTGATGAT 1183  
Db 1139 CAAAGTCACTAATATGCTTTTTCATGATCCAAAGAAATTAAGAAATTTGATGAT 1198  
Qy 1184 GGGAGGCTTCCCAACTTTCTTACGAAGATTTGCGAAATTTGAGGCTTATGTT 1243  
Db 1199 GGGAGGCTTCCCAACTTTCTTACGAAGATTTGCGAAATTTGAGGCTTATGTT 1258  
Qy 1244 TCTATTTGAAAAATATGTCAGATGAGAAATGATGTAATTAAGAGCAACAT 1303  
Db 1259 TTTATTTTGAAGAAAGCTGTAAG---AGCAATCTGATCAAGAAATATGAGAGACAT 1315  
Qy 1304 GTTAACAAAGAGCTTGAATGCAAGGCAATTAAGTCTGCAAGAAATATGAGAGAA 1363  
Db 1316 GCTAACAAAGAGCTTGAATGCAAGGCAATTAAGTCTGCAAGAAATATGAGAGAA 1375  
Qy 1364 TGTATCATGTTCAACAAAGCTGTTGAAAAGTTTCAACAGAAATTAAGATAT 1423  
Db 1376 TGTCTCATGTTCAACAAAGCTGTTGAAAAGTTTCAACAGAAATTAAGATAT 1435  
Qy 1424 GTGAGAACTGAAGAAAGAAAGATTTCTCAACGATGAATTAATTTCTGT 1483

```
Dh 1436 GCAAAAAAGTGAAGAAATTAAGAGTTTTCGAGTAAGATTAATTTGCTTAT 1495
Oy 1484 GTGTACAGCCAGCAAAAGCAGCCCGTTGCTTACATGATCTTGAATGAAGAACTATCT 1543
Db 1496 GTGTCAACATCAAAAGCAGCTGCAATTCCTTGCATGATCTTGCATGAAGAACTATCT 1555
Oy 1544 TTTTTCGACACACTGATCAAAAGCGAGATTTCCCGACAGATTAATAATTGCAAGGAT 1603
Db 1556 TTTTTCGACAGCATTTGAACGAAAGCAGATTTTCAGTGAAGAACATCGCGAAGGAT 1615
Oy 1604 TGGGAGAAAGTCCAGATTTAGAGAGGATTTCAAAAGAAATTAATGAGCGCATGCTATA 1663
Db 1616 TATTTAAAGAAATGTAGGCTCTTAAGAAAGATTTCTAAAAAAATTAATGCGCATGCTATA 1675
Oy 1664 CACTGAGCAGCAATGCAATCGCTTGGGAGCTACAGAAATTTTAAAGCAGGTTTATTGG 1723
Db 1676 CATTAAGCCAAAATTTGTGATCAATTTGAGAAACGTAAAGAAATGAAGAACTTTTACTAA 1735
Oy 1724 ATGAACACAAAGATCTTTGAAAGACAAAGAAAGTTGTATAAATTACTAAAGAAAGT 1783
Db 1736 ATGAACATAGAGATTAATTTGAAGAAATCAAGAAATTTGGAATGTATTGAAAGAGGAAT 1795
Oy 1784 GTAATTAATGGCTAGGAAGAGAGATGACCGTTTCTCTTTGTATGTGCTTCCAAAAG 1843
Db 1796 GCAATGAATGGCTAGGAAGAGAAATGAACGTTTCTCTTTTATGTGCTTTCGAAAAATA 1855
Oy 1844 CTACGTGTAGAGTGTATGTTAAAGACGTGAAGACAGGTGTGAATTTCAAAAAAAATA 1903
Db 1856 GGAATGAGAAATATGTGTGAAGATGTGAAGAAATCAATGCAAAATATTTGAAAAAACA 1915
Oy 1904 TAAAAAGCTTCATATATTTATGAATTTCTTGAAGAAATTAATCAATTAATTAACACACTGG 1963
Db 1916 TTTAAAAATACCAAGATATG-----ATAGTAAACCTAAATAGAAAGATTTAG 1963
Oy 1964 AAAGAAATTTGCTCTTGGCAGTACGTATGCAATAGATTTTCACTAATTTGCCAGGTC 2023
Db 1964 GGAATATTTGCTTATTTGGACCCACACTGCAATAGATTTGGACCAATTTGCCGAGATC 2023
Oy 2024 TTACGAAAGAGAAATGTGTACAAAAATCAAGAACATCGTAGCCGTTCTATTAAGAA 2083
Db 2024 TTGAAGAAA--AAATAATGTGAAGCTTTGAAAAAATTTGCAAACTTATTAATAGCAA 2080
Oy 2084 AGGCTTGAAGAGATCTCTCAAGTAGAGCTTCAAGAAATTTGACTGATTAATTTAAAT 2143
Db 2081 GAGACCTTGAAGAAATCAGCTTATTTAGTTAGTGAGGACATCTTGATTAAGAAAAAACT 2140
Oy 2144 GTGAACCTGCATTTGAAAGATATTGTACAGTAGCGGAAAGCTAATAATATGTCATATCA 2203
Db 2141 GCAAAACAAATCTTATTAAGTACTGTACATGTGGATCAACAGAAATTAACACACTTA 2200
Oy 2204 GTGGCTTATGCAAAAGCTAACAAGATTAATCTGTGAAGAGTAGTAGGATGCTAGAA 2263
Db 2201 AAGGTTTTTGTAA-----ACAGTTTCACTGATTAACATGAACATTATTAAG 2245
Oy 2264 AGGAATCTGTGAGAAATTTAGTGAAGAGTGAAGAAAGACAGTGCAGAACATTAACAACAG 2323
Db 2246 ATAAACTTTGGCAAAACCTAGTTCAAGCGTGTGAAGAAAAAGCCAGAGATTAATCAAAAG 2305
Oy 2324 AATTAGGACAACCGGAGCTGATCTAAAAAAAGATTATPAAGCATATGAGGAAGCTTAAGA 2383
Db 2306 AACTTGAAGAAAGCAAAATATTTAGAGAAACATTAAGATTAATGAAGAAAGTAAAG 2365
Oy 2384 AACGTGACAGAGAACAAATGAACAGTCCAGTCTTGTGTTGCTCACTCACTTAAGAAAAAG 2443
Db 2366 AGGATATCAAAAAATTCATATGAAGAAACAAATCTGTTTTTCAACATTAATCAACAG 2425
Oy 2444 AAAGTATATGTATCAAAAGTAAATAGCAAAAAACAGATPAAGATCCGTTCAACGAGAC 2503
Db 2426 ATAAATTAACAGAAAAAG-----AGTCAAGCCTAGTAGCGCTTACTGTAG 2470
Oy 2504 TTCAATGATACCAAAACATGTGAAGATTACTAGAGAGGGAAGTTAAGATGATATCCGTA 2563
Db 2471 TTCAAGATATTTGTACATTT-----AAACTGTAAAGAAAGAAATGAAGAAAGTTCAAGTGA 2524
```

```
Oy 2564 CAGAAATTAAGAGCTAAAGCATTTGATTTGGCAGAGAAATTAATTTGAAGATATGTAGATT 2623
Db 2525 CAGAAAAAGAAAGCAAAAGCGTTGATTTGGTAGACATGTACATGCTTATGTAGAGT 2584
Oy 2624 TGAAGAAAGATGTATTAATTAATTTGAATCAATTCGACAAATTAAGAGAGATTGCAAGACT 2683
Db 2585 TAAAAAGAAACGTGTACCATCTTAAGAGATGATTCGAATTTTGAAGAAAGATGTAAA--T 2641
Oy 2684 TAGAAGAGATGTGCAAAAAAGATTATAGAGCTTGTCCAAATCTGAAGCCTGTGAGAGTGA 2743
Db 2642 GTAAAGACCAATGCAAAAGAGATTAAGAAAAATATGTTTAAAAATTAAGAACATGAAAGTAA 2701
Oy 2744 AGCCGCAAGAAACGTGACAGAAAGTACAAAGCACTACAAACAACAACAAACCGTTG 2803
Db 2702 AGCCACATGAATTAATTAACAGTAAACGAAACCAACATPAACAACAGTCAACAAGCA 2761
Oy 2804 CCGATCCGAAAGCAACG-----AATGCAATTCCTTACAGCAACAGCA 2848
Db 2762 AAGAAGCAGAAAAACAGTAGAGAGACGAGAGAAATGCAAAATCTCAGCAACAAGACA 2821
Oy 2849 CATGGGTTTACACAGACATCGACACACAAGCAAGCTCTACTATGACATCTACATCAAT 2908
Db 2822 GTGGGTCAAAAGACGTCAACCCATACAGACTCCACAGACTACGTCACAGATTAAGT 2881
Oy 2909 CAAAAATPAACATTGACATCAACGAGGCGATGCAAAACCAACAGTGTACGACAGAGAGG 2968
Db 2882 CAAGATPAACACTGACCTCGACGAGGCGGTGTAAGCTTACGAAGTGTACGACAGAGAGG 2941
Oy 2969 AAGATGATGACGAGAGAGTGAACCGAGTAGGGGCTGAGAGATGAGTCGGTGAATGTGA 3028
Db 2942 AAGATGAAGCAGAGAGGTGAAGCCGAGTAGGGGCTGAGAGATGATGGGTGAAGTGA 3001
Oy 3029 TGAGGGGGGTGATAGTAGCAATGTTATTTGTTTCATGATTT 3069
Db 3002 TGAGAGGGGTGTTATTAAGCAATGATGATTTCAATTCATGATT 3042

RESULT 14
AR438435
LOCUS AR438435 3006 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6664053.
ACCESSION AR438435
VERSION AR438435.1 GI:42663290
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 3006)
AUTHORS Kovacs,J.A., Huang,S., Masbur,H., Fischer,S.H., Gyll,V.J. and Wei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
gene of human Pneumocystis carinii
JOURNAL Patent: US 6664053-A 3 16-DEC-2003;
FEATURES
Location/Qualifiers
1..3006
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 30.2%; Score 927.4; DB 6; Length 3006;
Beet Local Similarity 60.3%; Pred. No. 4,1e-149;
Matches 1859; Conservative 0; Mismatches 1131; Indels 93; Gaps 16;
Oy 2 TGGGCGGGCGGTCAAGCGGACAGGACAGAGATTAAGCATTTGATGAGGAACATG 61
Db 2 TGGGCGGGCGGTCAAGCGGCGGCTGCAG--CAAGAAATAGTGTGAAGAAATATC 58
Oy 62 TTTTACCTTAATTTCTAAAGAGATGAGCTAAGTAGACAGAAATGCAAAAAAACTAA 121
Db 59 TTTTGGCTTTGATTTTGAAGAAATGAGTAAATTAATGATTAATGTAAAAAAGTTGA 118
Oy 122 AAAAAATTTGCCAAGATTTGACTGAGCAAAACCTAAATATTAAGAACAGTACACAGAAAC 181
```

```

Db 119 AAGAGTATGTGAAGTTTAAAAAATGTAAACAAAAGAACCAAAAACTAGAAAGAAAGT 178
Qy 182 TTTAAAGTTTTTGGGAAGATGGAAGAAC--AGATACAAAATGCAAAAGAACTGAAGACCA 238
Db 179 TAGCGGAATCTGCAAGATGATATAACATAGAAAGCAAAATGCAAGAAACAGAAACAA 238
Qy 239 ATATTGAGAAAAAATGTACTACATCAAAAGGAAAACTTAAAGAAACAATTTAAAAA 298
Db 239 AGGTTAAAGCAAAAGTACTAGTTTTCAAACGAACTTGATTAAGGCACTCAAAAAGGAG 298
Qy 299 TTCAGATTATTAACGATTAAGGATTCGAAGAAGAAAGAAACAATGCCATTTTGGAG 358
Db 299 CTTCAACATTATAGAGATATGATGTAAAGAAATGAACGACAAATCCCTGTTTGGAG 358
Qy 359 GAGTATTTTCAAAAGAACTTAAAGATGATTCGCAATCTTTAGAAATTAAGTCATCAA 418
Db 359 GAGATGTCGAACGAACTTAAAGATTAATGAACTAGAGAAATTAATGTTATCAA 418
Qy 419 AGAAACGTGATTAAGTTGCGGAAGAAAGTTCTTTTAAGACACTTCGTAGCGATCTTAATG 478
Db 419 AAAAAACGAGACGATAGCAGAAAAAGCTCTTTTAAGGATCTTAAGAGGAACTTAAAG 478
Qy 479 GATCAGTCATATGTGAAGAAAAAATCTTAAAGAAATTTGCCCTGTCAAGGAGGAAAGTG 538
Db 479 ATAAAAACAATGCAAAAAATTAAGTTAAAGGGGTTGTCAAGAAATTCAAAGAAAGTG 538
Qy 539 ATGAGTTAACAACTTGTGTCTGAACCAAGAAAGACATGTAAATTTTAAATTTGAA 598
Db 539 ATGAGCTAAATTAATATGTCTTGAACGAAGAAAAACGTGTGAAGATCTTGTATCTAGA 598
Qy 599 AAGATTAAGAAAGTGGGATCTCTTAAACAGATGTTTCAGCAGCACTAGAA-----GTT 652
Db 599 AAGATTAACAAATGCAAACTCTCAAGAGAAATTTGATCTAGGCTTGGAAAGAAAGATT 658
Qy 653 TTTAAAAAGAAACGTCTTGAATTAATCTGAACAATGCTATTTTACATGGAATTCG 712
Db 659 TATTAAGAAAGAAATGTTATTTATTTCTGAAGAAATGTTACTTTTAAGGTCAAATCGTG 718
Qy 713 GAGACGAGATATTAATTAATGTATGAAATTTGGAGAGAAATGCGAAGAAACAAATATG 772
Db 719 AACAAGATCAACCC--AAGGTAAAGAGTTGCAAGCAATGTCAAAAGGAAATCTCG 775
Qy 773 CTTATATGCCACGAGACCCGATTTTGTATCCAACTAGGCCAGAGGCTACATATAGACAG 832
Db 776 TTTATGCAACGACCAAGGTTCAACATTTGATCTTAAGAAATTAAGATTAAGTTAGCAAG 835
Qy 833 ATATAGGCTGGAAGAGTTTATTAAGAGGTAGAGAGATGAGATTTTTAATTTGAAAGA 892
Db 836 AATATAGACCTAGAAAAATTTGTAAGTAAGAGAGTGAAGAAAGGAATTCATATGGAAGG 895
Qy 893 ATCATCTAAGAGATGCGACAGCTTTGTTGGCATTTGATCCAAAGATTTCTAGCTTAA 952
Db 896 CATCAATTAAGATGAGTCCGTTTATTTGCAATTAAGCA-----GAGTATGCTC 949
Qy 953 AAAAAAGCAAGAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGTCGAAAAATCTCATG 1012
Db 950 AAAATACTTTTAAAGATCAATGTGAAGATGTTATTAAAAAAATGTGGAACCTTTAAG 1009
Qy 1013 AACATGAGGCTTTAAGAAAGTTATGTAGAAAGAAATGTTTAAGTATGATGGAACGA 1072
Db 1010 AGCATATTTATTTAAAGATTTATGATGATTAAGATCAATCAATCAATCAAAAGAA 1069
Qy 1073 AATGTGAAGATTTGCAAAATGATATTAACAAACTTGGCAAAATTTTCACTTCAAAAGTCA 1132
Db 1070 AATCGAAGAACTAAATTAAGAGTT-----ACCAACCCGTAATTTTAACTGTTTCTAAA 1123
Qy 1133 CTAATTAATCGCTTTTGAATCAACAAAGAAATTAATGAATTTTGGATGGGAAGGTT 1192
Db 1124 GGATTTGAGAAATATTTGCTTCAAGCTAATGTAAA---GGAAATTAATTTGTTGGCAATGT 1180
Qy 1193 TGCCACAATTTCTTAGCAACGAAGATGTGCGAAATTTGAGTCTTATTTGTTTCTAATTTG 1252
Db 1181 TGCAATCAATTTCTTGTGGAAGAGAGTGTACGAACGTGTTGCGGATTTGTTTATTTGA 1240

```

QY 2333 AACCGGACGTGATCTATAAAAAAGTTATTAAGACTATGAGAACTTAAGAACTGCAG 2392  
DB 2279 AAGTAAAAAGAGTCTTATAGAAAAAGAAAGAAAGTATTAATAAATTAAGAAAGAGCAG 2338  
QY 2393 AGAAGCAATGAAACAAGTCCAGCTTGTGTTGTCTCATCTATTAAAGAAAAAGAAATATG 2452  
DB 2339 AAAAAAGCATGGAAGATGCAACCTTATTTATTCAGAGCCGAAAGAGACTGTATATAT- 2397  
QY 2453 TATCAAAAAGTAAATGCAAAAAACAAGATTAAGATGCCGTTTCAACCGACTTCAAGATA 2512  
DB 2398 -----AATTAATAAGTCAGTAAATAAAGACTCATCTGATACACCTTAAGAGAAAGCA 2452  
QY 2513 CCACAAAACATGTGAAATTAATCTACGAGGAGGAGTTAAGAGTATCCCTAACAGAAATAG 2572  
DB 2453 CAACGAGATTT-----AACTTGTAAAGAAAGAAAGCAAAAGTGCATTAACAGAAAAAG 2506  
QY 2573 AAGCTAAAGCAATTTGATTTGGCAGAGAAAGTATTGAAAGATATGTAGATTGAAAGAAA 2632  
DB 2507 AATTAGCAGCATTTATTTATTTGGTAGCAAGAGCATTTGATCTTATCTAATTAATTGAAAGAAA 2566  
QY 2633 GATGTAAATTAATTGGAATCAAGATTGAGATTGAGAGAGATTGCAAGACTTAAGAAAG 2692  
DB 2567 TATGTAAATCAATTCATGAAAGATTTGGTTTCAAAAAAGAGT--TGACTGTAGAGATC 2623  
QY 2633 TATGTAAAAAATTAATTAAGCTTGTCCCAATCTGAAGCTCTGAGAGTGAAGCCGACG 2752  
DB 2634 CATGTAAAAAGATACAGGAAATATGTTCAACATTAGAGCCACTTAAATTAAGTGAACACACG 2683  
QY 2753 AAACAGTGACAGAAAGTACAAAGCAACGACATCAACAAACA-----ACAAACAACGTTGCG 2806  
DB 2684 AATATGTAACTTAAATAACATTAACACTAACAACCAACCAACCACTACATTAAG 2743  
QY 2807 ATCCGAAGGCAACGGAATGCAATCTTTACAGACAAACAGACATGAGGTTACACAGACAT 2866  
DB 2744 ACGCAAGGCAACAGACTGCCACTTTACAGACAAACAGATACGTTGAGTCAAAAGAGT 2803  
QY 2867 CGACACACACAGACGCTTACTATCAACATCTACATCAATCAAAATAATCATTTGACAT 2926  
DB 2804 CGACCCATACATGACATCCACACACACATCACTACGTCACGTCAGAAATTAACGTTGACCT 2863  
QY 2927 CAACGAGGCGATGCAACCAACCAAGTGAACGACGAGAGAGAGATGATCAGAGAGACG 2986  
DB 2864 CGACAAAGCGGTGTAAGCTTACGAAGTGTACACAGAGAGAGATGAACAGAGACG 2923  
QY 2987 TGAACCGAGTGAAGGCGCTGAGAGTGAAGTGAAGTGAAGGAGGCGGTGATAGTAG 3046  
DB 2924 TGAACCGAGTGAAGGCGTTGAGAGTGAAGTGAAGTGAAGGAGGCGGTGATAGTAG 2983  
QY 3047 CAATGCTTATTTCTTCAATGATT 3069  
DB 2984 CAATGACGATTTCAATTCATGATT 3006

RESULT 15  
AF149015 3459 bp DNA linear PLN 20-JUL-2000  
LOCUS AF149015 3459 bp DNA linear PLN 20-JUL-2000  
DEFINITION Pneumocystis carinii f. sp. ratti major surface glycoprotein R7a  
(R7a) gene, partial cds, and major surface glycoprotein R7b  
gene, partial cds.  
ACCESSION AF149015  
VERSION AF149015.1 GI:9294725  
KEYWORDS Pneumocystis carinii f. sp. ratti  
SOURCE Pneumocystis carinii f. sp. ratti  
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 3459)  
AUTHORS Schaffzin,J.K. and Stringer,J.R.  
TITLE The major surface glycoprotein expression sites of two special  
forms of rat Pneumocystis carinii differ in structure  
JOURNAL J. Infect. Dis. 181 (5), 1729-1739 (2000)  
MEDLINE 20283743  
PUBMED 10823775

REFERENCE 2 (bases 1 to 3459)  
AUTHORS Schaffzin,J.K., Garbe,T.R. and Stringer,J.R.  
TITLE Direct Submision  
JOURNAL Submitted (06-MAY-1999) Molecular Genetics, Biochemistry, &  
Microbiology, University of Cincinnati College of Medicine, 231  
Bethesda Ave, Cincinnati, OH 45267-0524, USA  
FEATURES  
Source location/Qualifiers  
1..3459  
/organism="Pneumocystis carinii f. sp. ratti"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:38082"  
/note="forma\_specialis: ratti"  
1..500  
/gene="R7a"  
1..500  
/gene="R7a"  
/product="major surface glycoprotein R7a"  
1..500  
/gene="R7a"  
/codon\_start=3  
/product="major surface glycoprotein R7a"  
/protein\_id="AA06638.1"  
/db\_xref="GI:9294726"  
/translation="TTEETETVTEFEVLSPSTVPGKTEITTIKMTOTEIK  
YSSPIITTKRTIEGNTTICDINTQPTKTSFQITVTSVTTTIIISTYKLTSLK  
CTPKSTSDTKETQPEPTKSTDSIKETQKEEVAKNEGVKIRISEMTKIFFGIT  
IMGL"  
629..903  
/note="similar to the upstream conserved sequence (UCS)  
exon II"  
1..3459  
/gene="R7b"  
1..3459  
/gene="R7b"  
/product="major surface glycoprotein R7b"  
1..3459  
/gene="R7b"  
/note="silent donor MSG that in order to be expressed must  
be linked to a unique expression locus (UCS) whose coding  
sequence possesses an ATG start codon"  
/codon\_start=1  
/product="major surface glycoprotein R7b"  
/protein\_id="AA06639.1"  
/db\_xref="GI:9294727"  
/translation="AAQPVREAPKAPGPEPAVPGPGPSGIIQOVVAEGIGEE  
LILALIVKSYNNEGNCKTALKKYCDLANTDPTLQVNEKVYKLCCKDKEGKCKDL  
KGEVKKVGEFQKLEBALKKTGDEKCKLEBKCIPLSEADPTNLKENCYLNECY  
KBRKDVAREVLIKLVGVSVDGCKLBEVCKLSGVSDLMFCLDNNNTCNL  
NTRVNYCTPLKELKNSKEBEKLEDCYFSSSGDAKKKEIKQCEKEPTAT  
APSGFSPVERKPGFERKIGUEVLRBAEDGVLVGPGTLMKDLILLSKGETGV  
QKTCENVLVKNCDFRYLQELKGLGDTNGRKCKBELVADVSSRCSYLYKFLPK  
GLSTTFDGTSTNVLIMGELPTLISNEDCALESECKLESKCNKLKACGVAAC  
KKGODRYLVNLFEGRLBAHGSLSIGGOVCEVYIEKCKLEKREYVAKCLFPEGLC  
QOVLVEDI FVRLIEELGRVLDVDRPREDDCVLELGGKSHLGNVRLNEDKATSRRC  
HYLAVESQLKRDPLKRSVDVLSSEKKEBGLKERCODLNRRKRPESVSCALPETYC  
PFSNEIRIECIRFQDRVEDALKNGQOBOVCAPFAYTIDYGVCSITLIRKGNKDN  
GPFLEELKCGPFEELKLEBELLYELGSLTNSECKKLEKCTQPKNSINDTLTGL  
CKDNGNEBEVCKLVKVEKCKPSLNTNLEKAAVLEBKKEVDEYKKEKRPANTA  
SLVSRPKKEKPPAEPKPAPEKPAAGSAAANGITVKADVDASVYRIPTPR"

ORIGIN  
Query Match 13.0%; Score 399.8; DB 8; Length 3459;  
Best Local Similarity 51.6%; Pred. No. 1e-58;  
Matches 1242; Conservative 0; Mismatches 1062; Indels 105; Gaps 10;  
QY 44 GCATTGATGAGGAACATGTTTACCTTAATTTCAAGAAAGAGGACTAGAGACGAC 103  
DB 1022 GCATTGGGAGAACTTATTTGCTTTGATGGAAGAGTAAATTAATATGAAGCA 1081  
QY 104 AATCAAAAAAACTAAAAATATGCGCAAGATTGACTGAAGCAAAAACTAATATAG 163  
DB 1082 ATTGCAAAACAGCCCTCAAGAAATATTGTGATGATTGAATGAACAGATCGAACTTAC 1141

Qy 164 AACAGTACACAGAAAATTAAAGTTTTCGAGATGAGAAAAGCAGATACAAATGCA 223  
Db 1142 AGAAAGTGAATGAAAAGTTAAAGGACTTTGTAAGATTAAGACACAGAAAATTAATGCA 1201  
Qy 224 AAGAACTGAAAGCCCAATTGAGAAAATGTATCAATCAAGAAAATTAAAGAG 283  
Db 1302 AAGCTTGAAAGGAGAGTTGAAAAGTTGAGAGTTTAAACAGACCTTCAAGGTG 1261  
Qy 284 CAATTAAGAAAATTCAGATTTAAAGGATTAAGATTCGAAAGAGATGACAAAT 343  
Db 1262 CATTGAAAAAACAAGCGA-----TGAGAAAGATTTGTAAGAGACTTGAAGGAAAT 1312  
Qy 344 GCCATTTTTCGAGAGATATGTTCAAAAGAACTTAAAGATGATTCGAATCTTTGAGA 403  
Db 1313 GTATATTTTTCGAAAGAGCGAGTCCAACTTAATTTAAAGAGACGCGTAAGTTGAGA 1372  
Qy 404 ATAACTGCTATCAAAAGAAAGTGTAAAGTTGCGGAAGAGTTCTTTAAGACATTC 463  
Db 1373 ATGAATGCTATTAAGAAAGGAAAGATGTGCGAGAGAGTTCTTATTAAGAGTCTTA 1432  
Qy 464 GTACGATCTTAATGAGATCAGTATGTGAAAAAAAATTAAAGATTTGCCGTGCA 523  
Db 1433 AAGTTGTTGCGGATGAGTTGTAAGATTCGAAATGAGAAATGAGAGTTGTGCAAAAT 1492  
Qy 524 TGGGGAAGGAAAGTGAATGATTAACAACTTGTCTGACACAGAAAGACATGTAGA 583  
Db 1493 TAACTGATACAGGATGAATTTGATGTTAGTGTCTTGATCCAAATTAATGCGTCAAA 1552  
Qy 584 ATATTTTAATTGAAAAGATTAAGAGTCCGTACTTTAAACAGATGTTTCAAGACAC 643  
Db 1553 ATCT-----CAATACAGAGTGAACAAATGTTTGCAACCCCTTTAAAGAAAGAC 1600  
Qy 644 TAGAAGTTTAAAGAAAATCATGCTTGAATTAATTCGAAATGCAATGCTATTTTACATG 703  
Db 1601 TGAATAATGTAATCAAGAGAAATGTACGAAAAGCTTGAAGACTGCTATTTTACGAT 1660  
Qy 704 GAAATTCGAGAGACGATTAATTAATGTAATTTGAGTTGGAGAAATGCCAAGAC 763  
Db 1661 CAATGTTG-----CGATCAAAAAATGTAAGAAATTTGAAAGCAATGCCAGGAAA 1714  
Qy 764 AAAATATTGCTTAATATGACACAGAGCCGATTTTGAATCCAACTAAGCCAGGCTACAA 823  
Db 1715 AAGATTTACATACACCCGCGCAGAGTCCGTTTCACTCCGTTGMAAGAAACAGAGAT 1774  
Qy 824 TAGCAGAGATATAGGCGTGAAGAGTTTATTAAGAGTTAGAGAGATGAGAGTTTAA 883  
Db 1775 TCGAGAAAAGATTTGGTTGAAGAGTTCTTAAGAGACAGAAAGCAGATGAGGTTTGG 1834  
Qy 884 TTGGAAGAAATCATCTAAGAGATGCGACAGCTTTGTCATTTGATCCAAATCTA 943  
Db 1835 TTGGAAGACCGGATACAAAGATGCTGAAGATTTGATTCTGTTGTAGCAAAAGGGAGA 1894  
Qy 944 GTCTTAAAAAAAAGACGACAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAA 1003  
Db 1895 CTGCTAACGTTCA-----AAAAACAATAAGCAGAAATGTGTTAAAAAAAATTTGAT 1947  
Qy 1004 ATCTCATGAACAGAGGCTTTAGAAATTTATGTAAGAAAAATGCTTAAATGATATG 1063  
Db 1948 AAATTCAGATCTGAATCAGAG-----TTAAAGGGTTATGCGGTGATCAAAATGTA 2002  
Qy 1064 GAACGAAAAAATGTAAGATTTGCAAAATGATATTAAACAAATCTGCAAAATTTTCACTT 1123  
Db 2003 GACAAAABAAAATGTAAGATTTAGTAAAGATGTCTCATCCAGATGTTCAATCTTAAT 2062  
Qy 1124 CAAAAGTCACTAATATGCTCTTTTGTATCAACAAAAGAAATATGAAA--TTGTTG 1180  
Db 2063 ATAAGCTTTTTCGAAAGGGTGTCTAGACATTTGATGTAATAACGTAACCTTCTGT 2122  
Qy 1181 GATGGGAAGGTTGCAACATTTCTAGACAGAAATGTCGAATTTGAGAGTCCATTT 1240  
Db 2123 TATGGGAGACCTTCAACATTTGCTTACCAATGAAGATTTGCGAGAGCTAAGTCTGAT 2182  
Qy 1241 GTTCTATTTTGAATAAAATGTCGA--GATGAGAAATGATGTAAATATTAAGAG 1297

Db 2183 GTTATATCTTAGAGAGTTCAATGTAATAATCTAGAAAAAGCATGCAAAAATGTGAGG 2242  
Qy 1298 CAACATGTTCAAAAGAGCTTGATGACAGGCGAAATTAAGCTCAGCAAAAATATGAC 1357  
Db 2243 CAGCTTTATTAATAAGAGGACAGACAGGGTGTGAATGCAATGTTTGAAGAAAGTTAG 2302  
Qy 1358 GAGAAATGTTACATGTGTTCAAAACAAAAGCTGTTGAAAAGTTTCAACAAATTAAGTA 1417  
Db 2303 AAGCTCATGGAAT-----TGAGTTGCTTGATCAGAGG 2338  
Qy 1418 AAGTATGAGAAATCTGAAAAAGAAAAACAAGAGTTTCTCAACAGATGATTAATTA 1477  
Db 2339 AAGTGTGAAGTATGATGTAATAAATGTAAAGAACTTAAGATTAACAGAAATACCTG 2398  
Qy 1478 TTCTGTGTATACAGCCGCAAAAGCAGCCGGTTGCTTACACATGATCTCGAATGAAA 1537  
Db 2399 CAATAATCTTTTCCGGAGAAATATGTCAACAAAGTTGGAGAGATATTTTGTATGAT 2458  
Qy 1538 CTATCTTTTACGACAACTGATCAAAAAGCAGATTTCCGACAGATTAATAATGCA 1597  
Db 2459 TAGAGAAATTAAGAGAGTTTGTGAATGATGAGAGATTTTCCAAAGAGAGATATGCG 2518  
Qy 1598 AGGAATTTGGGAGAAAGTCCAGATTTAGAGAGATTCAAAAGAAATTAATGATGCCAT 1657  
Db 2519 TTGAATTTGGAAAGAAATGACAGCCATCTGGAATTAATTTGCGCTTGAATGAAGATTAAT 2578  
Qy 1658 GTCATATACGTGAGACCAATGCAATGCTTTGGGAGATACAGAAATTTAAAGAGTTT 1717  
Db 2579 GTGCTAATTAAGTGAAGCGCTGCTATTAATTAAGTGTGTTAGAGCAATGAGGAAAGATT 2638  
Qy 1718 TATGATGAAACAAAGATACTTTGAAGACCAAGAAAGTTGTGTAATAATACCTAAAG 1777  
Db 2639 TCTTAAAAAGAAACAGATGATTAATTAAGTGAAGAAAAATGTAAGAGAGTTGAAAAG 2698  
Qy 1778 AAAAGTAAATTAATGCTTGAAGAGAGATGACGTTTCTTTTGTATGTCTTCC 1837  
Db 2699 AAAAGTGTATCAACTGAATTAAGAGAGAAAGAACCCGTTAGTGTTCGCGCATATGC 2758  
Qy 1838 AAAACGCTAGCTGACATGATGTAAGACGTAAGAACAGCTGTGAATTTCAAAA 1897  
Db 2759 CAGAAAGAAACATGTAATTTTCTCAAAATGAATAAGATTAAGTGTCTTATATTAAG 2818  
Qy 1898 AAAATATTAAGCTTCATATATTAATTAATTTCTTGAATAATTAATCAATAATAACAA 1957  
Db 2819 ATAGGTTGAAGATGCTATGAAAAATTAAGACAAAAGATGACAACTTGCACCGCT 2878  
Qy 1958 CACTGAAAGAAATTTGCCCTTTGGCATGATTTGCAATAGATTTTCACTAATTTGTC 2017  
Db 2879 TCGCTCATATTAATGATGAATATGAAGAAAGTTTGTCAAGAAATCTT-----AGGAA 2929  
Qy 2018 CAGCTTTAAGAAAGAAATATGTTTACAAAATTAAGAAAGATCTGTAGCCGTTCTATA 2077  
Db 2930 AAGGTAATGTATTAAGATTAAGACCTTTTAAGAACTTGAAGAAATGTGTCCATCTTTG 2969  
Qy 2078 AAAGAAAGGCTTGGAAGATGCTCTCAAGTGAAGCTTCAAGGAAATTTGCTGATAAT 2137  
Db 2990 AAGCAATTAATTTGAAAGAGAGATTTATATGACTTAAGAGAGATTTAACCAATGATA 3049  
Qy 2138 CTAAATGTGAACCTGACATTAAGAAATATTTGTAACATGAGCGGAAACGTAAATTAATCGT 2197  
Db 3050 GTGAATGTGAAAAAATTTGAAGAGATATTGTAATCAAAATTTCAACAAATGATA 3109  
Qy 2198 CAATCAGTGTATGCAAGCTTAACCAAGATTAACCTGGAAGAGATGAGATG 2257  
Db 3110 CATTGACTGTATTAATTAAGATTAATA-----ATGGCG 3142  
Qy 2258 CTAGAAAGAACTGTGAGAAATTAAGTGAAGAGTGAAGAAACAGTGCAAAGCATTAAC 2317  
Db 3143 GGAATGAAGAGTTGCAAAAATTTAGTTAAAAAAGTGAAGAGAAATGTCTCTTTAA 3202  
Qy 2318 CAACGAATTAAGCAACCGGACGCTGATTAATAAAGATTAATAGACATATAGGAAC 2377

```
Db      3203 AAACCTAATCTGAAAAAGCGAAAGCTGAATTGGAAGAAAAAGAGGAATACGATGAGG 3262
OY      2378 TTAAGAAACGTGACAGAGGAGCAATGAACAAGTCCAGTCTTGTGTTGTCTCACTCATTAAGA 2437
Db      3263 TGAATAAAAGAGCGCAAAAAATTCCGGAATACAGCTAGCTGTGTGTGTGAGAGACCCAGAGA 3322
OY      2438 AAAACGAAA 2446
Db      3323 AAGGTGAGA 3331
```

Search completed: October 29, 2005, 06:20:32  
Job time : 12899 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 2005, 23:39:12 : Search time 1553 Seconds  
(without alignments)  
11709.881 Million cell updates/sec

Title: US-10-654-416-13

Perfect score: 3072  
Sequence: 1 atgcgcggcgccgcgcacgcg.....ttattcgttcacgatttag 3072

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_GeneSeq\_16Dec04:\*

- 1: geneSeqn1980s:\*
- 2: geneSeqn1990s:\*
- 3: geneSeqn2000s:\*
- 4: geneSeqn2001s:\*
- 5: geneSeqn2001bs:\*
- 6: geneSeqn2002as:\*
- 7: geneSeqn2002bs:\*
- 8: geneSeqn2003as:\*
- 9: geneSeqn2003bs:\*
- 10: geneSeqn2003cs:\*
- 11: geneSeqn2003ds:\*
- 12: geneSeqn2004as:\*
- 13: geneSeqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3072	100.0	3072	3	AA294069 Pneumocys
2	2318.8	75.5	3054	3	AA294068 Pneumocys
3	2262.6	73.7	3084	3	AA294066 Pneumocys
4	1900.6	61.9	3090	3	AA294065 Pneumocys
5	1624.8	52.9	3081	3	AA294067 Pneumocys
6	1098.6	35.8	3042	3	AA294063 Pneumocys
7	927.4	30.2	3006	3	AA294064 Pneumocys
8	289.6	9.4	2110	2	AAQ41223 Clone PC3
9	261.6	8.5	2814	2	AAQ40201 Sequence
10	238.8	7.8	3521	2	AAQ41226 Clone GP3
11	211	6.9	3521	2	AAQ41230 Gene enco
12	192	6.2	1448	2	AAQ40202 Sequence
13	191.2	6.2	1454	2	AAQ41224 Clone PCS
14	185	6.0	249	3	AA294070 P. carini
15	146.2	4.8	2190	2	AAQ41225 Clone PC1
16	137.2	4.5	2126	2	AAQ41225 Clone PC1
17	121.8	4.0	2058	2	AAQ41227 Clone GP4
18	91.4	3.0	1300	12	ADP85917 Synthetic
19	87	2.8	6292	4	AA546735 Tumour su
20	85.8	2.8	2110	2	AAQ41228 Clone GP2

21	83.4	2.7	1000	12	ADQ62833	Adg62833 Homopoly-
22	83.4	2.7	1000	12	ADQ62832	Adg62832 Homopoly-
23	79.2	2.6	6644	2	AA333181	Aax33181 Base sequ
24	79.2	2.6	7372	2	AA333182	Aax33182 Base sequ
25	79.2	2.6	7797	2	AA333180	Aax33180 Cowpox vi
26	79.2	2.6	7996	2	AA333184	Aax33184 Base sequ
27	78.2	2.5	9539	4	AA545347	Aa545347 Chemicall
28	78.2	2.5	9539	6	ABK28180	Abk28180 DNA trans
29	77	2.5	12237	6	ABL34358	Ab134358 Human imm
30	77	2.5	14006	6	ABL33958	Ab133958 Human imm
31	76.2	2.5	5750	4	AA546708	Aa546708 Tumour su
32	76.2	2.5	5750	6	ABL34008	Ab134008 Human imm
33	76	2.5	2520	12	ADJ48201	Adj48201 Maize oil
34	75.8	2.5	5181	6	ABL70443	Ab170443 Chemicall
35	75.8	2.5	5493	6	AA561386	Aa561386 Human gen
36	75.6	2.5	529	13	ACNS6029	Acns6029 Cotton an
37	75.6	2.5	627	13	ACNS4555	Acns4555 Cotton an
38	75.4	2.5	388	13	ACNS5013	Acns5013 Cotton no
39	75.4	2.5	700	10	ACD92384	Ac92384 Human col
40	75.4	2.5	3566	6	AA563367	Aa563367 Chemicall
41	75	2.4	525	11	ACN80333	Acn80333 Breast ca
42	75	2.4	6668	6	ABL33697	Ab133697 Human imm
43	75	2.4	16033	6	ABL33404	Ab133404 Human imm
44	74.8	2.4	489	13	ACNS4837	Acns4837 Cotton an
45	74.6	2.4	529	8	ABX52180	Abx52180 Bovine ES

## ALIGNMENTS

RESULT 1  
AA294069 ID AA294069 standard; DNA; 3072 BP.

AA294069; (revised)  
15-SEP-2003  
05-JUN-2000 (first entry)

Pneumocystis carinii major surface glycoprotein gene HMSG35.

Major surface glycoprotein; MSG; HMSG35; human; pneumonia; diagnosis; ds.

Pneumocystis carinii, sp. f. hominis.

OS XX  
FH XX  
FT XX  
FT CDS  
FT Key  
FT Location/Qualifiers  
FT 1..3072  
FT /\*tag= a  
FT /product= "HMSG33"  
FT /note= "a nucleic acid comprising residues 2821-3072 of  
this sequence is specifically claimed in Claim 28"

WC020009760-A1.  
24-FEB-2000.  
17-AUG-1999; 99MO-US018750.  
17-AUG-1998; 98US-0096805P.  
(USSH ) US DEPT HEALTH & HUMAN SERVICES.

Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;  
WPI; 2000-206025/18.  
P-PSDB; AAY79171.

Detection of the presence of Pneumocystis carinii in specimens by  
identification of major surface glycoprotein (MSG) gene sequences using  
two or more oligonucleotide primers derived from human P. carinii MSG  
protein encoding sequence.

Claim 27; Page 95-99; 110pp; English.



XX The present sequence is that of the novel *Pneumocystis carinii* sp. f.  
CC hominis gene, HMSG35, which encodes a major surface glycoprotein (MSG,  
CC see AAY791711). The gene was isolated by PCR amplification of DNA taken  
CC from an autopsy lung sample of an HIV-infected patient with *P. carinii*  
CC pneumonia. It is 1 of 7 novel, claimed genes (see AAY79063-69) of the  
CC invention that encode human-*P. carinii* MSGs (see AAY79165-71). The MSGs  
CC include a highly conserved C-terminal region of approximately 100 amino  
CC acids; this region (also claimed) corresponds to residues 2821-3072 of  
CC the present sequence. Direct detection or amplification of human-*P.*  
CC *carinii* MSG-encoding genes, especially by PCR using primers directed at  
CC the conserved region of the genes, provides a sensitive and specific  
CC technique for the detection of *P. carinii*, and the diagnosis of *P.*  
CC *carinii* pneumonia, especially in biological specimens (e.g. blood,  
CC sputum) from immunocompromised patients such as those with HIV infection.  
CC (updated on 15-SEP-2003 to standardise OS field)

Sequence 3072 BP; 1232 A; 438 C; 683 G; 719 T; 0 U; 0 Other;

Query Match 100.0%; Score 3072; DB 3; Length 3072;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3072; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGCGCGGCGGTCAAGCGGCGAGCGAGGAGACAGAAATGATGATGAGGAGAT 60
DB 1 ATGGCGCGGCGGTCAAGCGGCGAGCGAGGAGACAGAAATGATGATGAGGAGAT 60
QY 61 GTTTAGCTTTAATCTAAGAGATGAGCTAAGTGAGCAGAAATGCAAAAAAAGCTA 120
DB 61 GTTTAGCTTTAATCTAAGAGATGAGCTAAGTGAGCAGAAATGCAAAAAAAGCTA 120
QY 121 AAAAAATTTGGCAAGATTGCTGAGCAAACTAATATATGACAGTACAGAAA 180
DB 121 AAAAAATTTGGCAAGATTGCTGAGCAAACTAATATATGACAGTACAGAAA 180
QY 121 AAAAAATTTGGCAAGATTGCTGAGCAAACTAATATATGACAGTACAGAAA 180
DB 121 AAAAAATTTGGCAAGATTGCTGAGCAAACTAATATATGACAGTACAGAAA 180
QY 181 CTTAAGGTTTTTGGAAGATGAGAAAGCAGATACAAAATGCAAAAGAACTGAAGCCAT 240
DB 181 CTTAAGGTTTTTGGAAGATGAGAAAGCAGATACAAAATGCAAAAGAACTGAAGCCAT 240
QY 181 CTTAAGGTTTTTGGAAGATGAGAAAGCAGATACAAAATGCAAAAGAACTGAAGCCAT 240
DB 181 CTTAAGGTTTTTGGAAGATGAGAAAGCAGATACAAAATGCAAAAGAACTGAAGCCAT 240
QY 241 ATTGAGAAAAATGTACTACATCAAGAGAAAATTAAAGAAAGCAATTTAAAAAATTT 300
DB 241 ATTGAGAAAAATGTACTACATCAAGAGAAAATTAAAGAAAGCAATTTAAAAAATTT 300
QY 301 CAGATTATACCGATAGAGATTGCAAGAGAAATGAAACAATGCTATTTTGGAGGGA 360
DB 301 CAGATTATACCGATAGAGATTGCAAGAGAAATGAAACAATGCTATTTTGGAGGGA 360
QY 361 GATGTTCAAAAAGAACTTAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420
DB 361 GATGTTCAAAAAGAACTTAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420
QY 421 AAACGTGATAAAGTTGCGGAAGAAAGTTCTTTTAAAGACCTTCGTAGCGATCTTAATGA 480
DB 421 AAACGTGATAAAGTTGCGGAAGAAAGTTCTTTTAAAGACCTTCGTAGCGATCTTAATGA 480
QY 481 TCAGTCATATGTGAAAAAACTTAAGAGATTTCCTGTCATGCGGAGGGAAGTAT 540
DB 481 TCAGTCATATGTGAAAAAACTTAAGAGATTTCCTGTCATGCGGAGGGAAGTAT 540
QY 541 GAGTTTACAAACTTGTGTCGAAACAGAAAGAGATGTAAGATTTTAAATTTGA AAAA 600
DB 541 GAGTTTACAAACTTGTGTCGAAACAGAAAGAGATGTAAGATTTTAAATTTGA AAAA 600
QY 601 GATAAGAGTGCGGTAATCTTTAAACAGATGTTTTCAGACGACTAGAAATTTAAAAA 660
DB 601 GATAAGAGTGCGGTAATCTTTAAACAGATGTTTTCAGACGACTAGAAATTTAAAAA 660
QY 661 GAAACATGCTTGAATTAATCTGAAACAATGCTATTTTACATTTGAAATTTGCGAGACAC 720
DB 661 GAAACATGCTTGAATTAATCTGAAACAATGCTATTTTACATTTGAAATTTGCGAGACAC 720
QY 721 GATATATTAATTAATGATTTGAAATTTGGAGGAAAAATGCAAGAAATTAATGCTTATATG 780
DB 721 GATATATTAATTAATGATTTGAAATTTGGAGGAAAAATGCAAGAAATTAATGCTTATATG 780
```

```
DB 721 GATATATTAATTAATGATTTGAAATTTGGAGGAAAAATGCAAGAAATTAATGCTTATATG 780
QY 781 CCACGAGACCCGATTTTGTATCCAACTAGAGCCAGAGGCTACAAATGACAGAGATATAGG 840
DB 781 CCACGAGACCCGATTTTGTATCCAACTAGAGCCAGAGGCTACAAATGACAGAGATATAGG 840
QY 841 CTGGAAGAGTTTATTAAGAGGTAGAGAGATGGAAGTTTATTTGCAAAATCACTTA 900
DB 841 CTGGAAGAGTTTATTAAGAGGTAGAGAGATGGAAGTTTATTTGCAAAATCACTTA 900
QY 901 AGAGATGCGACAGCTTTGTTGGCAATTTGATCCAAATTTAGTCTTAATTAATTAATTAAT 960
DB 901 AGAGATGCGACAGCTTTGTTGGCAATTTGATCCAAATTTAGTCTTAATTAATTAATTAAT 960
QY 961 GACAAAGAAATGCGAAGAAAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAACATGAG 1020
DB 961 GACAAAGAAATGCGAAGAAAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAACATGAG 1020
QY 1021 GCTTTAGAAAATTTATGTAAGAAAAATGCTTTAATGTAATGATGGAACGAAAAATTTGTA 1080
DB 1021 GCTTTAGAAAATTTATGTAAGAAAAATGCTTTAATGTAATGATGGAACGAAAAATTTGTA 1080
QY 1081 GAATTCGAAAATGATATTTAACAAAATCTGCAAAATTTTCACTTCAAAAAGTCACTAATAT 1140
DB 1081 GAATTCGAAAATGATATTTAACAAAATCTGCAAAATTTTCACTTCAAAAAGTCACTAATAT 1140
QY 1141 CGTCTTTTGTATCCAAAGAAATTAATGAATTTGTTGATGAGAGAGGTTGCCAACA 1200
DB 1141 CGTCTTTTGTATCCAAAGAAATTAATGAATTTGTTGATGAGAGAGGTTGCCAACA 1200
QY 1201 TTTCTTAGCAACGAAGATTTGCGCAATTTGAGATCTATTTCTATTTTGAATTAATTA 1260
DB 1201 TTTCTTAGCAACGAAGATTTGCGCAATTTGAGATCTATTTCTATTTTGAATTAATTA 1260
QY 1261 TGTCGAGATGAGAAATGCAATGTAATAATTAAGACAAATGATGTTCAAAAAGAGACTT 1320
DB 1261 TGTCGAGATGAGAAATGCAATGTAATAATTAAGACAAATGATGTTCAAAAAGAGACTT 1320
QY 1321 GATCAGCGGCAATTAAGTCTGCAAGAAATTAAGCAAGAAATGTAATGATGTTCAAAAC 1380
DB 1321 GATCAGCGGCAATTAAGTCTGCAAGAAATTAAGCAAGAAATGTAATGATGTTCAAAAC 1380
QY 1381 AAAAGCTGCTTGAAGATTTTCAACAGAAATTAATTAAGATTAATTAATTAATTAATTA 1440
DB 1381 AAAAGCTGCTTGAAGATTTTCAACAGAAATTAATTAAGATTAATTAATTAATTAATTA 1440
QY 1441 GAAAAAGAAAGATTTCTCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
DB 1441 GAAAAAGAAAGATTTCTCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
QY 1501 GCAGCCCGGTTGCTTACATGATCTTTCGAATGAAAACTATCTTTTTCGACCAACATG 1560
DB 1501 GCAGCCCGGTTGCTTACATGATCTTTCGAATGAAAACTATCTTTTTCGACCAACATG 1560
QY 1561 GATCAAAAAGCAAGATTTTCCGACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
DB 1561 GATCAAAAAGCAAGATTTTCCGACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
QY 1621 GATTTAGAGAGATTTCAAAAAGAAATTAATGAGCATGTCATACACTGAGACAGCAATGC 1680
DB 1621 GATTTAGAGAGATTTCAAAAAGAAATTAATGAGCATGTCATACACTGAGACAGCAATGC 1680
QY 1681 AATGCTTTGGGAGTACAGAAATTTTAAAGCAGGTTTATTTGATGTAACACAAAAGATAT 1740
DB 1681 AATGCTTTGGGAGTACAGAAATTTTAAAGCAGGTTTATTTGATGTAACACAAAAGATAT 1740
QY 1741 TTGAAAGACCAAGAAAGTTGTGTAATAATCTTAATAAGAAAGTGTATTAATTTGCTTGA 1800
DB 1741 TTGAAAGACCAAGAAAGTTGTGTAATAATCTTAATAAGAAAGTGTATTAATTTGCTTGA 1800
QY 1801 AGAGAGATGACGCTTTCTCTTTTGTATGTTCTTCCAAAAGCGTACGATGAGCTGATG 1860
DB 1801 AGAGAGATGACGCTTTCTCTTTTGTATGTTCTTCCAAAAGCGTACGATGAGCTGATG 1860
```



Qy	1861	TTAAAGA	CGGAAAGACAGGGTGGAGTAATTTCAAAA	AAAAATATPAAAGCTCATPATTT	1920
Dp	1861	GTAAAA	GACGTGAAAGACAGGTGTGAGATATTTCAAAA	AAAAATATPAAAGCTCATPATTT	1920
Qy	1921	ATTGAATTTCTT	GAATAATATATACAAATATAATTAACACACTG	GAAGAAATTTGCTCCCTCT	1980
Dp	1921	ATTGAATTTCTT	GAATAATATATATACAAATATAATTAACACACTG	GAAGAAATTTGCTCCCTCT	1980
Qy	1981	TGGCATTCG	TATTCGCAATPAGATTTTCACTTAATTTGTC	CAAGGTCTTACGAAGACATATGCT	2040
Dp	1981	TGGCATTCG	TATTCGCAATPAGATTTTCACTTAATTTGTC	CAAGGTCTTACGAAGACATATGCT	2040
Qy	2041	TGTACAAA	AAAAATCAAGAACATCTGTGAGCCGTTCTTA	TAAAGAAAGGCGCTTGGAAGATGCT	2100
Dp	2041	TGTACAAA	AAAAATCAAGAACATCTGTGAGCCGTTCTTA	TAAAGAAAGGCGCTTGGAAGATGCT	2100
Qy	2101	CTCAAA	AGTAGCTTCAAGGAAAATTTGACTGATTAATCT	TAAATGTGAACCTGTGATTTGAA	2160
Dp	2101	CTCAAA	AGTAGAGTACGCTTCAAGGAAAATTTGACTGAT	TAAATGTGAACCTGTGATTTGAA	2160
Qy	2161	AGATATTG	TGTAACGATAGCGGGAACGCTAATATATG	TGTCATCAGTGGCTTATGCAAAAGCT	2220
Dp	2161	AGATATTG	TGTAACGATAGCGGGAACGCTAATATATG	TGTCATCAGTGGCTTATGCAAAAGCT	2220
Qy	2221	AACACCA	AGATAACTCTGTGAAAAGATGATGAGATGCT	TAGAAAGAACTCTGTGAGAAA	2280
Dp	2221	AACACCA	AGATAACTCTGTGAAAAGATGATGAGATGCT	TAGAAAGAACTCTGTGAGAAA	2280
Qy	2281	TTTACTG	AAAGAAAGTGTGAGAAACAATTAACCAACGA	ATTAGCAACCGGCA	2340
Dp	2281	TTTACTG	AAAGAAAGTGTGAGAAACAATTAACCAACGA	ATTAGCAACCGGCA	2340
Qy	2341	GCTGATCT	TAAAAAAAGATTTATPAGACATATAGAGAA	CTTAAGAAACGTGCAGAGGAACA	2400
Dp	2341	GCTGATCT	TAAAAAAAGATTTATPAGACATATAGAGAA	CTTAAGAAACGTGCAGAGGAACA	2400
Qy	2401	ATGAA	CAAGTCCAGTCTTTGTTTGTCACTCATTA	AGAAAACGAAAGTATGTATCAAA	2460
Dp	2401	ATGAA	CAAGTCCAGTCTTTGTTTGTCACTCATTA	AGAAAACGAAAGTATGTATCAAA	2460
Qy	2461	AGTAA	TAGCAAAATCTGAGTCTTTGTTTGTCACTCAT	TAAAGAAAACGAAAGTATGTATCAAA	2520
Dp	2461	AGTAA	TAGCAAAATCTGAGTCTTTGTTTGTCACTCAT	TAAAGAAAACGAAAGTATGTATCAAA	2520
Qy	2521	CATGTGA	AAATACTACGAGGGGAGTTAAGGATGATAT	CCGTAAACGAATTAAGAACTTAA	2580
Dp	2521	CATGTGA	AAATACTACGAGGGGAGTTAAGGATGATAT	CCGTAAACGAATTAAGAACTTAA	2580
Qy	2581	GCATT	TGATTTGGCGACGAGAAAGTATTTGAAAGAT	TATGTAATTTGAAAGAAAGTATAT	2640
Dp	2581	GCATT	TGATTTGGCGACGAGAAAGTATTTGAAAGAT	TATGTAATTTGAAAGAAAGTATAT	2640
Qy	2641	AAAT	TGGAATCAGATTGCGAGAAATTAAGAGATT	TGCAAAAGCTTAAGAAAGTATGCAAA	2700
Dp	2641	AAAT	TGGAATCAGATTGCGAGAAATTAAGAGATT	TGCAAAAGCTTAAGAAAGTATGCAAA	2700
Qy	2701	AAGATTA	ATAAGGCTTTGTTCGCAATCTGAAGCCTCT	GAGGTGAAGCCCAAGAAACAGTG	2760
Dp	2701	AAGATTA	ATAAGGCTTTGTTCGCAATCTGAAGCCTCT	GAGGTGAAGCCCAAGAAACAGTG	2760
Qy	2761	ACAGAA	ATACAGCAACCTAACAACAACAACACG	CTTGCCGATTCGAAAGCAACG	2820
Dp	2761	ACAGAA	ATACAGCAACCTAACAACAACAACACG	CTTGCCGATTCGAAAGCAACG	2820
Qy	2821	GAAT	GCAAAATCCTTACAGACAAACAGACAT	GGGTTTACACAGACATCGACACACAAGC	2880
Dp	2821	GAAT	GCAAAATCCTTACAGACAAACAGACAT	GGGTTTACACAGACATCGACACACAAGC	2880
Qy	2881	ACGCT	TATATCAACATCTACACATCACTCA	AAAAATTAACATTGACATCAACAGAGGAGTGC	2940
Dp	2881	ACGCT	TATATCAACATCTACACATCACTCA	AAAAATTAACATTGACATCAACAGAGGAGTGC	2940

QY	2241	AAACCAACCAAGTGTACGACGAGGAGAAATGATGTCAGAGACGTGAACCGAGTGGAG	3000
Db	2941	AAACCAACCAAGTGTACGACGAGGAGAAATGATGTCAGAGACGTGAACCGAGTGGAG	3000
QY	3001	GAGCTGAGATGTAGTGGGTGGAATGTATGATGAGGGGGGTGATAGTGGCAATGTTATTTCG	3060
Db	3001	GCGGTGAGATGTAGTGGGTGGAATGTATGATGAGGGGGGTGATAGTGGCAATGTTATTTCG	3060
QY	3061	TTTCATGATTTAG 3072	
Db	3061	TTTCATGATTTAG 3072	
RESULT 2			
AAZ94068			
ID	AAZ94068	standard; DNA; 3054 BP.	
AC			
XX	AAZ94068;		
XX			
DT	15-SEP-2003	(revised)	
DT	05-JUN-2000	(first entry)	
XX			
XX			
XX			
XX			
OS			
XX	Major surface glycoprotein; MSG; HMSG33; human; pneumonia; diagnosis; ds.		
XX	Pneumocystis carinii; sp. f. hominis.		
Key	Location/Qualifiers		
FT	1..3054		
FT	/*tag= a		
FT	/product= "HMSG33"		
FT	/note= "a nucleic acid comprising residues 2887-3132 of		
FT	this sequence is specifically claimed in Claim 28"		
XX			
XX	MO200009760-A1.		
XX			
PD	24-FEB-2000.		
XX			
PF	17-AUG-1999; 99WO-US018750.		
XX			
PR	17-AUG-1998; 98US-0096805P.		
XX			
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
PI	Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;		
DR	WPI; 2000-206025/18.		
DR	P-PSDB; AAY79170.		
PT	Detection of the presence of Pneumocystis carinii in specimens by		
PT	identification of major surface glycoprotein (MSG) gene sequences using		
PT	two or more oligonucleotide primers derived from human P. carinii MSG		
PT	protein encoding sequence.		
XX			
PS	Claim 27; Page 87-91; 110pp; English.		
XX			
CC	The present sequence is that of the novel Pneumocystis carinii sp. f.		
CC	hominis gene, HMSG33, which encodes a major surface glycoprotein (MSG,		
CC	see AAY79170). The gene was isolated by PCR amplification of DNA taken		
CC	from an autopsy lung sample of an HIV-infected patient with P. carinii		
CC	pneumonia. It is 1 of 7 novel, claimed genes (see AAY79063-69) of the		
CC	invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs		
CC	include a highly conserved C-terminal region of approximately 100 amino		
CC	acids) this region (also claimed) corresponds to residues 2887-3132 of		
CC	the present sequence. Direct detection or amplification of human-P.		
CC	carinii MSG-encoding genes, especially by PCR using primers directed at		
CC	the conserved region of the genes, provides a sensitive and specific		
CC	technique for the detection of P. carinii, and the diagnosis of P.		
CC	carinii pneumonia, especially in biological specimens (e.g. blood,		
CC	sputum) from immunocompromised patients such as those with HIV infection.		
CC	(Updated on 15-SEP-2003 to standardize OS field)		
XX			

Sequence 3054 BP; 1222 A; 449 C; 670 G; 713 T; 0 U; 0 Other;

Query Match 75.5%; Score 2318.8; DB 3; Length 3054;

Best Local Similarity 86.0%; Pred. No. 0;

Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;

```
OY 4 GGGGGGGGGCTCAACGGCAGS---CAGAGGAGACACGAATGCAATTGATGAGAGAACT 60
DB 1 GGGGGGGGGCTCAACGGCAGGTAACAGAGCATCAGGGCAATATGATGATGAACTGAT 60
OY 61 GTTTTAGCTTTAATCTAAGAGAAATGCACTAAGCTAGCAGAGATGCAAAAAAACTA 120
DB 61 ATTTTGGCGTTGATCTACAGAGAGATGCAATGAGAGATGCAAAATGCAAAAAAGTTTA 120
OY 121 AAAAAATATGGCAAGATTTGACTGAGCAAAACTAATATAGAACAAAGTACACAGAAA 180
DB 121 GAAAAATATCTGCGAAGATTTGAAAAAGCATCTAGACATGAAAAAGTACATATAATG 180
OY 121 CTTAAGGTTTTTGGGAAGTGGAAAAAGCATACAAAATGCAAAAGAACTGAAAGCAAT 240
DB 121 CTTAAGGTTTTCTGGAATGGAAGCAAGCAAGTAAAGCAAAATGCAAAATGTCAGGCTCA 240
OY 241 ATTGAGAAAAAATGTACTACATCAAGAGAAAACTTAAAGACAAATTTAAAAAAATTT 300
DB 241 CAAGCCAAAGTTACGGGGAAATGTACAAATTTTAAACACAAAGCTAGACACGCTTA 300
OY 301 CAGATTTATACGAGATPAGGATTCGAAAGAAATGACAAATGCTATTTTGGAGGGA 360
DB 301 ACAATCCATCGATGATATTTGCAAGAGAGTGAACGACAAATGCTATTTTGGAGGGA 360
OY 361 GTATGTTCAAAAGACTTAAGATGATTGCAATCTTTGAGAAATPAGGCTATCAAAAG 420
DB 361 GATG---CCATPACTTTGAGAAATGTATCAAACTAAGGAATCTATGTAACAGAAA 417
OY 421 AAACGTGATTAAGTTGCGGAGAAAGTTCTTTTAAGACACTTCGATGCGATCTTATGSA 480
DB 418 AAACGTGACGAGATGACAGAGAAAGTCTTTTGAAGGACCTTCGATGATCTCAATAAA 477
OY 481 TCAGTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTCAATGGAGGAAAGTAT 540
DB 478 ACAGAAACACATGAAAAAACTGAAAGAGATTTCCCGAGCTCTCAGAGGGAAGTAT 537
OY 541 GAGTTAACAACTGTGCTGTAACAGAAAGACATGTAAGATATTTTAATGAAAAA 600
DB 538 GAATTTACGAGACTGTGTGTTGAACAGAAAGACGTGCAAGAAATTTAATAAAAAA 597
OY 601 GATTAAGAGTGGCTACTCTTAAACAGATGTTTCAAGCAGACTAGAGATTTTAAAAA 660
DB 598 GATTAAGAGTGGCTACTCTTAAACAGATGTTTCAAGCAGACTAGAGATTTTAAAAA 657
OY 661 GAAACATGCTTTGAAATTAACGACATGCTAATTTTCAATTTGAAATTTGCGAGACAC 720
DB 658 GAAATATGCTTTGAAATTAACGACATGCTAATTTTCAATTTGAAATTTGCGAGACAC 717
OY 721 GATTAATTAATTAATGATTAATTTGGAGGAAAAATGCAAGAACAAATTTGCTATATG 780
DB 718 GATTAATTAATTAATGATTAATTTGGAGGAAAAATGCAAGAACAAATTTGCTATATG 777
OY 781 CCACACAGACCCGATTTTGTATCCAACTAGGCCAGAGGCTACATATAGCAGAGATTA 840
DB 778 CCACACAGACCCGATTTTGTATCCAACTAGGCCAGAGGCTACATAGCAGAGACCTTA 837
OY 841 CTGGAAGAGTTTATTAAGAGGTAGAGAGATGAGATTTTATTTGAAAGATCATCTA 900
DB 838 CTGGAAGAGTTTATTAAGAGGTAGAGAGATGAGATTTTATTTGAAAGATCATCTA 897
OY 901 AAGAGATGCAAGCTTTGTGTCATTTG---ATCCAAAGATTTAGTCTTAAAAAA 957
DB 898 AAGAGATGCAAGCTTTATTTGAGCTTTGTTTGAAGAAAGATGATACAGAAAAAATAT 957
OY 958 GACGACAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAAC 1017
DB 958 AATATCGAGAAATGCAATGAATTTCTGAAAGATTAATGCAAAAATCTCTAAGACAT 1017
```

```
OY 1018 GAGCTTTAGAAAATTTATGTAAGAAAAATGCTTTAAGTAATGATGACAGAAAAATGCT 1077
DB 1018 GAACTCTTAAAAAATTTATGTAATTAATTAATGCTTATGATGATGAAAAAGAAAAAGC 1077
OY 1078 GAAAGATTTGCAAAATGATATTAACAAACTGCAAAATTTTCACTTCAAAAGTCACTAT 1137
DB 1078 AAGAAATTTAGAGAGATTTAATAAATATGATGACAAACCTCAAAACACAGATCTTAA 1137
OY 1138 AATGCTCTTTTGTATCCAAACAAAGAAATATGAAATTTGTGATGGAAGGGTTGCCA 1197
DB 1138 AATCATCTTTATGATCC-----AATGATAAATTTGTGATGGAAGAAAACTGCCA 1188
OY 1198 ACATTTCTTACGACAGAAATTTGCGAAATTTGAGTCTTATTTGTTCTATTTTGA 1257
DB 1189 ACATTTCTTACGACAGAAATTTGCGAAATTTGAGTCTTATTTGTTCTATTTTGA 1248
OY 1258 AATGTCAGATGAGAAAAATGCAATGTAATAATTAAGAGCAATGTTACAAAGAGGA 1317
DB 1249 GCTTGTCAAAATGCAAAAGAGATGTAATTTGAGGCGAGCTGTTATTAAGAGAGG 1308
OY 1318 CTTGATGACGGGCAATTAAGTCTGCAAGAAATATGCGAGAAATGTTACATGTTCA 1377
DB 1309 CTTGATGAGCGGGCAATTAAGTCTGCAAGAAATATGCGAGAAATGTTACATGTTCA 1368
OY 1378 AACAAAGCTGGCTTGAAGATTTCAACAGAAATTAATTAAGATGTAAGAACTGAAA 1437
DB 1369 AATCAAGTTGGCTTGAAGATTTCAACAGAAATTAATTAAGATGTAAGAACTGAAA 1425
OY 1438 AAGAAAAAAGAAAGATTTCTCAAGATGTAATTTATTTCTGTGCTACAGCAGCA 1497
DB 1426 AAGAAAAAAGAAAGATTTCTCAAGATGTAATTTATTTCTGTGCTACAGCAGCA 1485
OY 1498 AAAGCAGCCGGTCTTACATGATCTTGAATGAAACATATCTTTTTCAGCAACAA 1557
DB 1486 AAAGCAGCAGATTTACTTACACAGATCATCAATGAGGTTATCTTTTTCAGCAACAA 1545
OY 1558 CTGATCAAAAGCGAATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGAGAAAGTGC 1617
DB 1546 TTGATCAAAAGAGATTTTCCGACAGATTAAGATGCAAGAAATTTGGAGAAAGTGC 1605
OY 1618 CAAGATTAAGAGAGATTTCAAAAGAAATTTACATGCGCATGCTATACCTGAGACAGAA 1677
DB 1606 CAAGATTAAGAGAGATTTCAAAAGAAATTTACATGCGCATGCTATACCTGAGACAGAA 1665
OY 1678 TGCAATGCTTTGGGACTACAGAAATTTTAAGCAGGTTTATTTGATGAACACAAAGAT 1737
DB 1666 TGCAATGCTTTGGGACTACAGAAATTTTAAGCAGGTTTATTTGATGAACACAAAGAT 1725
OY 1738 ACTTTGAAGACCAAGAAATTTGTATAAATCTTAATAAGAAAGTGTATAATGCTCT 1797
DB 1726 ACTTTGAAGACCAAGAAATTTGTATAAATCTTAATAAGAAAGTGTATAATGCTCT 1785
OY 1798 AGAAGAGAGATGACCGTTTCTTTTGTATGTGTCTTCCAAACGCTACGATGAGCTG 1857
DB 1786 AGAAGAGAGATGACCGTTTCTTTTGTATGTGTCTTCCAAACGCTACGATGAGCTG 1845
OY 1858 ATGTTAAAGACGTGAAGACAGGTGTGAATTAACAAAAAATATAAAGCTTCATAT 1917
DB 1846 ATGTTAAAGACGTGAAGACAGGTGTGAATTAACAAAAAATATAAAGCTTCATAT 1905
OY 1918 ATTAATGAATTTCTTGAATAATTAATCAATTAATAACACATGGAAGAAATTTGCTCC 1977
DB 1906 ATTAATGAATTTCTTGAATAATTAATCAATTAATAACACATGGAAGAAATTTGCTCC 1965
OY 1978 TCTTGGCATAGATTAATGATATTTTCACTAATTTGCTCAGGCTTACAGAAAGAT 2037
DB 1966 TCTTGGCATAGATTAATGATATTTTCACTAATTTGCTCAGGCTTACAGAAAGAT 2025
OY 2038 AGTTGTACAAAAATCAAGAAACATGTGAGCCGTTCTATAAAGAAAGCCCTTGAGAT 2097
DB 2026 AGTTGTACAAAAATCAAGAAACATGTGAGCCGTTCTATAAAGAAAGCCCTTGAGAT 2085
```

QY 2098 GCTCTCAAGTAGAGCTTCAAGAAAAATGACTATTAATCTAAATGTGAACCTGCATTG 2157  
 DB 2086 GCTCTCAAGTAGAGCTTCAAGAAAAATGACTATTAATCTAAATGTGAACCTGCATTG 2145  
 QY 2158 AAAAGATATTTGTACAGTAGCGGGAACGTAATAATGCGTCAATCAGTGGCTTAATGCAA 2217  
 DB 2146 AAAAGATATTTGTACAGTAGCGGGAACGTAATAATGCGTCAATCAGTGGCTTAATGCAA 2205  
 QY 2218 GCTTAACCAAGAGTAACCTCTGGAAGAGTAGAGATGCTTAAGAAAGAACTGTGAG 2277  
 DB 2206 GCTTAACCAAGAGTAACCTCTGGAAGAGTAGAGATGCTTAAGAAAGAACTGTGAG 2265  
 QY 2278 AAATTAGTAAAGAGTAGAGAAACAGTCAACCAATTACCAAGAAATTAGCAACCG 2337  
 DB 2266 AAATTAGTAAAGAGTAGAGAAACAGTCAACCAATTACCAAGAAATTAGCAACCG 2325  
 QY 2338 GCAGCTGATCTAAAAAAGATTATPAAGCATATGAGAACTTAAGAAACGTGCAGAGAA 2397  
 DB 2326 GCAGCTGATCTAAAAAAGATTATPAAGCATATGAGAACTTAAGAAACGTGCAGAGAA 2385  
 QY 2398 GCATTAAGCAAGTCCAGCTCTTTTGTGCACTCACTTAAGAAAAAGAAATGTATCA 2457  
 DB 2386 GCATTAAGCAAGTCCAGCTCTTTTGTGCACTCACTTAAGAAAAAGAAATGTATCA 2445  
 QY 2458 AAAAGTAAAGCAAAAAACAGATPAAGATCCGTTTCAACGCACTTCAAGATACACA 2517  
 DB 2446 AAAAGTAAAGCAAAAAACAGATPAAGATCCGTTTCAACGCACTTCAAGATACACA 2505  
 QY 2518 AAACATGTGAATAACTACGAGAGGAGATTAAAGATGTATCCGTAAACGAATTAGAACT 2577  
 DB 2506 AAACATGTGAATAACTACGAGAGGAGATTAAAGATGTATCCGTAAACGAATTAGAACT 2565  
 QY 2578 AAACATTTGATTTGGCAGCAGAAATATTGGAAGATGTATGATTTGAGAAAGATGT 2637  
 DB 2566 AAACATTTGATTTGGCAGCAGAAATATTGGAAGATGTATGATTTGAGAAAGATGT 2625  
 QY 2638 AATAAATTTGATGATGAGATTGAGAAATTGAGAGATTGCAAAAGCTTGAAGAATGTGC 2697  
 DB 2626 AATAAATTTGATGATGAGATTGAGAAATTGAGAGATTGCAAAAGCTTGAAGAATGTGC 2685  
 QY 2698 AAAAAGATTAATAAGCTTTGTGCAATCTGAAGCCTTGGAGGTGAACCCGCAAGAAACA 2757  
 DB 2686 AAAAAGATTAATAAGCTTTGTGCAATCTGAAGCCTTGGAGGTGAACCCGCAAGAAACA 2745  
 QY 2758 GTGACAGAAAGTACAGCAACATTCACACACACACACACACACACACACACACACAC 2817  
 DB 2746 GTGACAGAAAGTACAGCAACATTCACACACACACACACACACACACACACACACAC 2805  
 QY 2818 ACGGAATGCAAAATCCTTACAGACACACACACACACACACACACACACACACACACAC 2877  
 DB 2806 ACGGAATGCAAAATCCTTACAGACACACACACACACACACACACACACACACACACAC 2865  
 QY 2878 AGCAGCTCTACTACATCTACATCAATCAAAAATTAACATTGACATCAACGAGCGGA 2937  
 DB 2866 AGCAGCTCTACTACATCTACATCAATCAAAAATTAACATTGACATCAACGAGCGGA 2925  
 QY 2938 TGCAAAACCAACCAAGTGTACAGACAGAGAGAAATGATGACAGAGAGCTGAACCGAGT 2997  
 DB 2926 TGCAAAACCAACCAAGTGTACAGAG-----GGATGATGACAGAGAGCTGAACCGAGT 2979  
 QY 2998 GAGGGGCTGAGAGTAGAGGTGAGATGTGATGAGGGGGTGAATGATGACATGGTATT 3057  
 DB 2980 GAGGGGCTGAGAGTAGAGGTGAGATGTGATGAGGGGGTGAATGATGACATGGTATT 3039  
 QY 3058 TCGTTTCATGATTTAG 3072  
 DB 3040 TCGTTTCATGATTTAG 3054

RESULT 3  
 AA294066  
 ID AA294066 standard; DNA; 3084 BP.  
 XX

AC AA294066;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Pneumocystis carinii major surface glycoprotein gene HMSG14.  
 XX  
 KM Major surface glycoprotein; MSG; HMSG14; human; pneumonia; diagnosis; ds.  
 XX  
 OS Pneumocystis carinii; sp. f. hominis.  
 XX  
 FH Key  
 FT CDS  
 FT 1. 3084  
 FT /tag= a  
 FT /product= "HMSG14"  
 FT /note= "a nucleic acid comprising residues 2839-3084 of  
 FT this sequence is specifically claimed in Claim 28"  
 XX  
 PN W0200009760-A1.  
 PD 24-FEB-2000.  
 XX  
 PF 17-AUG-1999; 99WO-US018750.  
 XX  
 PR 17-AUG-1998; 98US-0096805P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;  
 XX WPI; 2000-206025/18.  
 DR P-PSDB; AAY79168.  
 XX  
 PT Detection of the presence of Pneumocystis carinii in specimens by  
 PT identification of major surface glycoprotein (MSG) gene sequences using  
 PT two or more oligonucleotide primers derived from human P. carinii MSG  
 PT protein encoding sequence.  
 XX  
 PS Claim 27; Page 72-76; 110pp; English.  
 XX  
 CC The present sequence is that of the novel Pneumocystis carinii sp. f.  
 CC hominis gene, HMSG14, which encodes a major surface glycoprotein (MSG,  
 CC from AAY79168). The gene was isolated by PCR amplification of DNA taken  
 CC from an autopsy lung sample of an HIV-infected patient with P. carinii  
 CC pneumonia. It is 1 of 7 novel, claimed genes (see AA279063-69) of the  
 CC invention that encode human P. carinii MSGs (see AAY9165-71). The MSGs  
 CC include a highly conserved C-terminal region of approximately 100 amino  
 CC acids; this region (also claimed) corresponds to residues 2839-3084 of  
 CC the present sequence. Direct detection or amplification of human P.  
 CC carinii MSG-encoding genes, especially by PCR using primers directed at  
 CC the conserved region of the genes, provides a sensitive and specific  
 CC technique for the detection of P. carinii, and the diagnosis of P.  
 CC carinii pneumonia, especially in biological specimens (e.g. blood,  
 CC sputum) from immunocompromised patients such as those with HIV infection.  
 CC (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 3084 BP; 1240 A; 455 C; 676 G; 713 T; 0 U; 0 Other;  
 Query Match 73.7%; Score 2262.6; DB 3; Length 3084;  
 Best Local Similarity 84.7%; Pred. No. 0;  
 Matches 2640; Conservative 0; Mismatches 399; Indels 78; Gaps 6;  
 QY 1 ATGGCGGCGCGGTCAAGCGGCAAGGACACAGCAATATGATGAGAAACAT 60  
 DB 1 ATGGCGGCGCGGTCAAGCGGCAAGGACACAGCAATATGATGAGAAACAT 60  
 QY 61 GTTTAGCTTAATTTAAAGAGATGAGCTAAGTACAGCAATGCAAAAACTA 120  
 DB 61 GTTTAGCTTAATTTAAAGAGATGAGCTAAGTACCAAAATGCAAAATCAAACTAAGT 120  
 QY 121 AAAAAATATTTGCCAAGATGAGTGAAGCAAACTAAATATGAAACAAGTACAGAAAA 180  
 DB 121 GAAAGATATTTGCCAAGATGAGTGAAGCAAACTAAATATGAAACAAGTACAGAAAA 180

Oy	181	CTTAAAGSTTTTTCGGAAATGGAAAAGCAGATACAAATGCAAAAGAACTGAAAGCCAAAT	240
Db	181	TTAAAGATTTCTGGATPACGGGAAAGAAATGAAAATGCAAGATCTAAAAACAA	240
Oy	241	ATTGAGAAAAATGATCTACATCAAAAGAAAACCTTAAAGAAAGCAATTAAAAAAATTT	300
Db	241	GTCATATCAAAATATGATTTAAATTTTCAGGAAAACCTTCAACAGCTGCTAGAAAAAAATTT	300
Oy	301	CAGATTTATPACGATPAGGATTTGCAGAAAGATGAAACAACATATGCTATTTTTCGAGGA	360
Db	301	TCAGATTTAACGATGAGGATTTGCAGAAAAGATGAAACAACATATGCTATTTTTCGAGGA	360
Oy	361	GTAATGTTCAAAAGAACTTAAAGATGATTTGCAATCTTTGAGAAATTAAGTCTATCAAAAG	420
Db	361	GCATGTCCAAACAGAACTTAAAGATGATCTGCAATTAATTAAGAAATTAATCTGTTATCAAAA	420
Oy	421	AAACGTGATTAAGTTCGGGAGAAAGTTCTTTTAAAGACATCTCGAGGCATCTTAATGA	480
Db	421	GACGGAACAAATGTGGCAGAGAAAGTTCTTTTGAAGGCGCTTCGTGTGATCTCAATGA	480
Oy	481	TCAGTCATATGTGAAAAAAAACCTTAAAGATTTGCCCCCTGTCATGCGGAGGAAAATGAT	540
Db	481	ACAAAGACATGTGAAAAAAAACCTGAAAGAAAGTTGCCGAAATTGAAAGAGAAAGCGAT	540
Oy	541	GAGTTAACAACTTGCTGTCTGAACGAGAAAGAGATGTAAAGATATTTTAATGAAAA	600
Db	541	GAATTTAACGAGCTTGTGCTTTATCAAAAAACATATCGTAAAGCTTGTAAACAAAGGA	600
Oy	601	GATAAGAAAGTCGCGACTCTTTPAAAACAGATGTTTCAGAGACACTGGAAGTTTTPAA---	657
Db	601	AAAAGTAATGTGATCTCTTGAAAAGAAAGTTGAAGAAAGCACTTAAAGAAATGATTTG	660
Oy	658	AAAGAAACATGTCTTGAAATTACTCGAACATATCTATTTTTCATTTGAAATTCGGAG--	715
Db	661	CGAGAAAATGTCTACTATTACTTGAGCAATGTTACTTTCACAGAGGAAACGTGAAAGGA	720
Oy	716	-----ACGACGATATAATTAAGT	735
Db	721	GACAAATCAAAAGTCGAATPAACTTAATATAATAAGCTGCAAAAGAAATTAATGACAGAGGT	780
Oy	736	ATTGAATTTGGAGGAAAAATGCCAGAAACAAATATTTGCTTAATATGCCACAGAGCCGAT	795
Db	781	GATGAATTTAGCAGAAAAGTGGAAAAAGAAATATTTGTTTAATATGCAATCCAGATCCGAT	840
Oy	796	TTTGATCCAACTAGGCCCAAGGCTTCAATAGCAGAGATATAGGGCTTGAAGAGTTTAT	855
Db	841	TTTCGATCCAACTAACGCCAGAGCTTACACTAGAGAGGACATATAGGGCTTGAAGAGCTTAT	900
Oy	856	AAGAAGATPAGAGGAGATGGAGTTTATTTGGAAAGATCAATCTPAGAGATGCGACACT	915
Db	901	AAGAGGCGAAGAGAGATGGAATTTTGTGGAAAGCAATGTAAAGATGCAACACT	960
Oy	916	TTGTTGGCATTTGATCCAAAGTTCTAGTCTTAAAAAAAAGACGACAAAGAGAAATGC	975
Db	961	TTGTTGGCACTACT-----TCTTAAGAAAACCCCTTAABAAAAGAAAGATGT	1005
Oy	976	GAAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAACATGAGGCTTTTGAAGTTTA	1035
Db	1006	ATPAAAGCCCTTAAAAAAAACCTGCAAAAACCTCATGAACATGAGGCTTTGAAAAATCTA	1065
Oy	1036	TGTAAAGAAAAATGGTTAAGTAAATGATGAGAAAGAAAAATGTAAGAAATTCGAAAAATAT	1095
Db	1066	TGTAAAGAAAAATPAAACAAATGATGATGAGAAAGAAAAATGTAAGAACTGAAAAAGAT	1125
Oy	1096	ATTAACAAACTGTGCAAAATTTTCACTTCAAAAGTCACATAATATCGTCTTTTGTGATCA	1155
Db	1126	GTTAACAAACTTGTACAAAGTCTTACATCAACAAATTTCTTAAAAACCGTCTTTTACATTTCA	1185
Oy	1156	ACAAAGAAATATATGAATTTGTTGATGAGGAAAGGCTTGCCAACTTTCTTATGCAACGA	1215
Db	1186	CC-----TGATGAATATGCGGAATAGGGGAAAAATTACCGACATTTCTTATGATGATA	1236

QY	1216	GATTGTGCGAAATTTGGAGTCTTATGTTTCTATTTTGA	1275
Db	1237	GATTGTGCGAAATCTAGAAATCTTATGCTTTTATTA	1296
QY	1276	AATGTCATGTAAAAATATAAGAGCAACATGTTTACAAA	1335
Db	1297	GAACTTGTATGAATGTGAAGGCAAGCTGTTCAGAAGAG	1356
QY	1336	AAAGTGTGCAAGAAAATATGCGAAGATGTACATGGT	1395
Db	1357	AGTGTGTGCAAAAATATATGCGGTGTATTTACGTGTT	1416
QY	1396	AAAGTTTCAACAGAATTAAGTAAAGTATGTGAAGAA	1455
Db	1417	GAGTTTCAACAAAGATTAGTAAAGTATGTAGAGAGT	1473
QY	1456	TTCTCAAAACGATGAATATTTATTCGTGTGTACAGC	1515
Db	1474	TTCCCAACGATGAATATTTGTTCTGTGTGTACAGC	1533
QY	1516	ACACATGATCTTTCGAATGAAAACCTATCTTTTTCGA	1575
Db	1534	ACACACATCATCAAAATGAGGGTACTCTTTTTCAGCA	1593
QY	1576	TTCCCGACAGATAAAAATTCGAAGGAAATGGGGAA	1635
Db	1594	TTTCCGACAGATAAAGACTCGCAAGGAATGCGGAA	1653
QY	1636	TCAAAACAAAATTAATGCGCATGTCACTGAGACAG	1695
Db	1654	TCAAAAGAAATTAATATGGCATGTCACTGAGACAG	1713
QY	1696	ACAGAAATTTTAAAGCAGTTTTTATGATGAAACA	1755
Db	1714	ACAGAAATTTTAAAGCAGTTTTTATGATGAAACA	1773
QY	1756	AGTTGTGTAAATATCTTAAAGAAAAGTGTATATAT	1815
Db	1774	AGTTGTGTAAATATCTTAAAGAAAAGTGTATATAT	1833
QY	1816	TTCTCTTTTGTATGTCTTCCAAAACGTCACGTGAG	1875
Db	1834	TTCTCTTTTGTATGTCTTCCAAAACGTCACGTGAG	1893
QY	1876	GACAGGTGTGAAGTATCAAAAAATATATAAAAGCT	1935
Db	1894	GACAGGTGTGAAGTATCAAAAAATATATAAAAGCT	1953
QY	1936	AATATATCAAAATAAATTAACAACACTGGAAAAG	1995
Db	1954	AATATATCAAAATAAATTAACAACACTGGAAAAG	2013
QY	1996	AATATGATTTTCACTTAATTTGTCACAGTCTTA	2055
Db	2014	AATATGATTTTCACTTAATTTGTCACAGGCTTA	2073
QY	2056	AAGATGCTGAGCCGTTCTATTAAGAAAGACCT	2115
Db	2074	AAGATGCTGAGCCGTTCTATTAAGAAAGACCT	2133
QY	2116	CAAGCAAAATTTGACTGATTAATCTTAATGTGA	2175
Db	2134	CAAGCAAAATTTGACTGATTAATCTTAATGTGA	2193
QY	2176	GCGGGAACGTAAATTAATGCGTCAATCAGTGGCT	2235
Db	2194	GCGGGAACGTAAATTAATGCGTCAATCAGTGGCT	2253
QY	2236	TTCTGGAAGAGTGTGAGATGCTGAAGAAAGAA	2295
Db	2254	TTCTGGAAGAGTGTGAGATGCTGAAGAAAGAA	2313
QY	2296	GAAAGAACAGTGCAGAAAGCATACCAACAGAA	2355

Accession	Gene	Chromosome	Position (kb)	Sequence	Length (bp)
Db	2214	GAAAGAACGTGCCAAGCACTTACCAACAGAAATTAGACAAACCGGACGCTGACTTAAAAA	2373		
Qy	2256	GATTATTAAGACATATGAGGAACTTAAAGAAACGTGACAGAGAACCAATGAACGTCAGT	2415		
Db	2274	GATTATTAAGACATATGAGGAACTTAAAGAAACGTGACAGAGAACCAATGAACGTCAGT	2433		
Qy	2416	CTTGTTTTTGTACCTCATTTAAGAAAAACCAAAAGTATGTATCTAAAAAGTAAATGCAAAAAC	2475		
Db	2434	CTTGTTTTTGTACCTCATTTAAGAAAAACCAAAAGTATGTATCTAAAAAGTAAATGCAAAAAC	2493		
Qy	2476	AAGGATTAAGATGCGCTTTCAAACGGACTTCAAGATACACAAAACATGTGAAATACTA	2535		
Db	2494	AAGGATTAAGATGCGCTTTCAAACGGACTTCAAGATACACAAAACATGTGAAATACTA	2553		
Qy	2536	CGGAGGGGAGTTAAGATGTAATCCGTAAACAGAAATTTAGAACCTAAAGCATTTGATTTGGCA	2595		
Db	2554	CGGAGGGGAGTTAAGATGTAATCCGTAAACAGAAATTTAGAACCTAAAGCATTTGATTTGGCA	2613		
Qy	2596	GCAGAAATTTTGGAAAGATATGTAGATTTGAAGGAAAGATGTAATTAATTTGGAATCAGAT	2655		
Db	2614	GCAGAAATTTTGGAAAGATATGTAGATTTGAAGGAAAGATGTAATTAATTTGGAATCAGAT	2673		
Qy	2656	TGCAGAAATTTAAGAGATGTCGAAAGACTTAAAGAAAGATGCAAAAGATTAATPAAGGCT	2715		
Db	2674	TGCAGAAATTTAAGAGATGTCGAAAGACTTAAAGAAAGATGCAAAAGATTAATPAAGGCT	2733		
Qy	2716	TGTGCGCAATCTGAAAGCCCTTGGAGGTGAAGCCGACGAAACATGTACAGAAAGTACAAAG	2775		
Db	2734	TGTGCGCAATCTGAAAGCCCTTGGAGGTGAAGCCGACGAAACATGTACAGAAAGTACAAAG	2793		
Qy	2776	ACAACATCAACACAAACAAACCGGTCGCAATCCGAATCCGAAGCGAATGCAATCTCTTA	2835		
Db	2794	ACAACATCAACACAAACAAACCGGTCGCAATCCGAATCCGAAGCGAATGCAATCTCTTA	2853		
Qy	2836	CAGACAACAGACACATGGGTTACACAGACATCGACACACAAAGACGTCATCATATCAACA	2895		
Db	2854	CAGACAACAGACACATGGGTTACACAGACATCGACACACAAAGACGTCATCATATCAACA	2913		
Qy	2896	TCTACATCAACATCAAAATAACATTGACATCAACGAGGCGATCAACCAACCAAGTGT	2955		
Db	2914	TCTACATCAACATCAAAATAACATTGACATCAACGAGGCGATCAACCAACCAAGTGT	2973		
Qy	2956	ACGACAGAGAGGAAGATGATGTCAGAGACGCTGAAAACCAAGTGAAGGGCTGAGATGAGT	3015		
Db	2974	ACGACAGAGAGGAAGATGATGTCAGAGACGCTGAAAACCAAGTGAAGGGCTGAGATGAGT	3027		
Qy	3016	GGGTGGAATGTGATGAGGGGGGTGATAGTACCAATGCTTATTTGTTTCAATGATTTAG	3072		
Db	3028	GGGTGGAATGTGATGAGGGGGGTGATAGTACCAATGCTTATTTGTTTCAATGATTTAG	3084		

XX WO200009760-A1.  
 XX  
 PN 24-FEB-2000.  
 PD  
 XX  
 XX 17-AUG-1999; 99WO-US018750.  
 PF  
 XX 17-AUG-1998; 98US-0096805P.  
 PR  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 XX Kovacs JA, Huang S, Maaur H, Fischer SH, Gill VJ, Mei Q;  
 PI  
 XX WPI; 2000-206025/18.  
 DR P-PSDB; AAY79167.  
 XX  
 XX  
 PT Detection of the presence of *Pneumocystis carinii* in specimens by  
 PT identification of major surface glycoprotein (MSG) gene sequences using  
 PT two or more oligonucleotide primers derived from human *P. carinii* MSG  
 PT protein encoding sequence.

CC The present sequence is that of the novel Pneumocystis carinii sp. f.  
CC hominis gene, HMSG11, which encodes a major surface glycoprotein (MSG,  
CC see AAY79167). The gene was isolated by PCR amplification of DNA taken  
CC from an autopsy lung sample of an HIV-infected patient with P. carinii  
CC pneumonia. It is 1 of 7 novel, cloned genes (see AAY79063-69) of the  
CC invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs  
CC include a highly conserved C-terminal region of approximately 100 amino  
CC acids; this region (also claimed) corresponds to residues 2845-3090 of  
CC the present sequence. Direct detection or amplification of human-P.  
CC carinii MSG-encoding genes, especially by PCR using primers directed at  
CC the conserved region of the genes, provides a sensitive and specific  
CC technique for the detection of P. carinii, and the diagnosis of P.  
CC carinii pneumonia, especially in biological specimens (e.g. blood,  
CC sputum) from immunocompromised patients such as those with HIV infection.  
CC (Updated on 15-SEP-2003 to standardise OS field)

**SQ** Sequence 3090 BP; 1255 A; 447 C; 662 G; 726 T; 0 U; 0 Other;

Query Match	61.9%	Score 1900.6	DB 3	Length 3090
Best Local Similarly	77.8%	Pred. No. 0		
Matches 2430; Conservative	0	Mismatches	609	Indels 84
				Gaps 8

Qy 1 ATGGCGGCGGTCAAGCGGCAGGCAGCAGGACACAGAATAGCATTGATGAGGAACAT 60

Db 1 ATGGCGGGGGCTCAAGCGGGCGCAAAAGTGCACAGATAGCATTGATGAGGACAT 60

61 GTTTAGCTTAATTCTAAGGAGATGACTAAGTGAGCAGGAATGCAGAAAAAACTA 127

Db 61 GTTTAGCTTTGATTTTAAAAAATGATTAGAGATACAATAAGCTG 12

121 AAAAATATTTGCCAAGATTGACTGAAGCAAATAATATAGAACCAAGTACACAGAAA 18

Db 121 GAAGATATTGCCAAACATTAAACAATGCAGATTAAATCCAGAAAAAGTTCAACGAAAAA 18

181 CTTAAGGTTTTCGAAGATGGAAGCAGATACAAATGCAAGAACTGAAGCCAT 2

DB 181 TTAAAGATTTCTCTGATACGGGAACGAAATGAAATGTCAAGATCTAAATAACAA 2

QY 241 A I T G A A A A A A I G I A C I A C A A I C A A G G A A A A C I T A A A G A A G C A A I T A A A A A A A A A A A A I I 30

[illegible]

— — — — —

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Oy	421	AAAGCTGATAAGTTCGGAGGAAGTCTTTTAAGACACTTCGTACGATCTTAATGGA	480
Db	421	GAAAGGAAACAATGAGGACGAAGAAGTCTTTTGAGGGGCTTCGTGATCTCAATGAA	480
Oy	481	TCAGTCATATGTGAAAAAAAACCTTAAGAAGATTGGCCCTGTCAATGGAGGAAATGAT	540
Db	481	ACAAAGACATGTGAAAAAAAACCTGAAGAATTTGCCGAATTTAAGAAAGGAAGGAT	540
Oy	541	GAGTTAACAACTTGTGTCTGAACCGAAGAGACATGTAAGAATTTTATTTAAGAAAA	600
Db	541	GAATTAACGAGACTTGTCTTTATCAAAAAACAATGCTGAATCTTTGTAACAAAAGGA	600
Oy	601	GATAAGAAAGTCGGACCTCTTAACACGATTTTCAGACAGACATGAAGATTTTAA---	657
Db	601	AAAAGTAATGTGATCTCTTGAAAAAGAAATGAAGACCTTAAGAAAGAAATTTG	660
Oy	658	AAAGAAACATGTCTTGAATTAAGTCAACATGCTATTTTTCATTGGAAATATGGGAG--	715
Db	661	CGAGAAAAATGTCTATCTTTACTTTAGACGATGTTACTTTCACAGAGGAACTGTGAAGGA	720
Oy	716	-----ACGACGATATATTAAATGT	735
Db	721	GACAAATCAAGTGCATATTAACCTAATTAAGACATGCAGAAAGATATGTACAGAGTGT	780
Oy	736	ATTGAATTTGGAGGAAAAATGCCAAGACAAATATTGCTTATATGCCACCGACCCGAT	795
Db	781	GATGATATTAGACAGAAAGATGTGGAAAAGAAATATTTGTTATATGATCTCCAGATCCGAT	840
Oy	796	TTTGATCCAACTAGGCGCAGAGGCTACATATAGAGAGATATAGGGCTGTGAAGATTTTAT	855
Db	841	TTTCATCCAACTTAAGCCAGAGCTTACATACAGAGGACATAGGGCTGTGAAGACTTTAT	900
Oy	856	AAGAAGGTAGAAGAGATGAGATTTTATTTGAAAAGAAATCATCTPAAGATGCGACACT	915
Db	901	AAGAGGGCAGAAAGGATGGAATTTTGTGTGAAGACAAATGTATAGAGATGCAACGCT	960
Oy	916	TTGTGTGCATTTGTGATCCAAAGATTTCTAGTCTTAAAAAAAAGCGACAAAGAAATGC	975
Db	961	TTGTGTGCACACT-----TCTTAAGAAAACCTTAAAAAAAGAAATGT	1005
Oy	976	GAAAGAGCCTTCAAAAAAGCTGCAAAAATCTCATGAACATGAGGCTTATGAAGTTTA	1035
Db	1006	ATTAAGGCCCTTAAAAAAAACCTGCGAAAACCTCTGAACATGAGGCTTATGAAGATCTA	1065
Oy	1036	TGTAAAGAAAAATGTTTAAGTATGATGGAACGAAAAAATGTGAAGATTGCCAAAAATGAT	1095
Db	1066	TGTAAAGAAAAATTAACCAAGTATGATGGAACGAAAAAATGTGATGAACATGAAGAAAGAT	1125
Oy	1096	ATTAAACAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATTAATCGTCTTTTGATCCA	1155
Db	1126	GTAAACAACAACTTGTACAAAGTCTTACATCAACAATTTCTTAAAAAACCGCTTTTACATTCA	1185
Oy	1156	ACAAAGGAATTAATGAAATTTGTTGAGAGGGAAGGGTGGCCAATTTCTTAGCAAGAA	1215
Db	1186	CC-----TGATGGAATTCGGAAATGGGGAATTAACCGACATTTCTTAGATGAA	1236
Oy	1216	GATTGTGCGAATTTGGAAGTCCATATGTTTCTATTTTGAAAAAAATGTCAGATGAGAA	1275
Db	1237	GATTGTGCGAATCTAGAAATCTTATTTGCTTTTATTAAGAAACCTGTCCAGATGTCAA	1296
Oy	1276	AATGCATTAATAATATPAGACCAATGTTTACAAAGAGACCTTGATGACGGGCAAT	1335
Db	1297	GAAAGCTGTATGAATGTAGAGGCGCGGTTATTAAGAGAGGGCTTGATGACGGGGCAAC	1356
Oy	1336	AAAGTCTCGCAAGAAATATGCGAGGAATGTTACATGTTTCAACAAAGTGGCTTGAA	1395
Db	1357	AGTGTGTGCAAAAAAATATGCGAGGTTATTTGCTCATTAATTAAGATTTGGCTTAAG	1416
Oy	1396	AAAGTTCAACAAGAAATTTAGTAAAGTATGTAGAAACGAAAAAAAAGAAACAAAGAGAT	1455
Db	1417	AAATTTCAACAAGAAATTTAGCAAAAGTATGTGAGAACTG---AAAGAAATTAAGAGAGAT	1473
Oy	1456	TTCTCAACGATGAATTAATTAATCTGTGTGATACAGCCAGCAAAAGCACCGGCTGTCT	1515

Db	1474	TTCTCGAACCATGAAATTGTTTGTCTCTGTGATACACACGACCAAGCAGACCATTA	CTT	1533
OY	1516	ACACATGATCTTGGAAATGAAAACTATCTTTTACGACAACTCGATCAAAAGCGAGAT		1575
Db	1534	ACACATCACCATCAAAATGAGAGTTATCTTTTACGACAACTCGATCAAAAGAGAGAT		1593
OY	1576	TTCCCGCAGATTAATAATTCGACAGAAATTGGGGAGAAAGTCCAAAGTTTATGAGAGAT		1635
Db	1594	TTTCCCGCAGATTAAGACTGCAAGGAAATTGAGGAGAAAATGCCAAGTTTATGAGAAAGAT		1653
OY	1636	TCAAAAAATAATTCATGGCCATGTCATACACTGGAGCAGCAATGCAATTCGCTTGGGAGCT		1695
Db	1654	TCAAAAAATAATTACATGGCCATGTCATACCTGAGACAGCAATGCAATTCGCTTGGGAGAT		1713
OY	1696	ACAGAAATTTTAAAGCAGGTTTTTATTTGATGAACACAAAGATACTTTGAAAGACCAAGAA		1755
Db	1714	ACAGAAATTTTAAAGCAGATTTTTATTTGATGAACACAAAGATACTTTGAAAGATGATGA		1773
OY	1756	AGTTGTGTAATAATTCCTAAAGAAAAGTGTAATTAATGTGCTTGAAGAGAGATGACCGT		1815
Db	1774	AACTGTCAAAATATTTTAAAGAAAATGCCATAAAGTGCTTGAAAGGGGTGATGATCGT		1833
OY	1816	TTCTCTTTGTAATGTGCTTCCAAAAGCTACGCTGAGCTGATGTGTAAGAAAGACGTGAA		1875
Db	1834	TTTCTCTTTGTAATGTGTTTTTCCAAAAGCTACATGTGAGCTGATGTAAAGAGCTGCA		1893
OY	1876	GACAGGTGTGAAGATTTCAAAAAAATAATAAAGCTTCATATAATTATGAAATTTCTTGAA		1935
Db	1894	GATAGGTGCAAAATATTCGAAGAAAATATGCAAGCATCAGATATTAATGATTTCCCTTAA		1953
OY	1936	AATATATCAAAATTAATAATPAACACTGGAAAAGAAATTTGCTCTTGGCATACGATATTC		1995
Db	1954	AAAAATCAAAATTAAGAGCAGATCAGCAGCAAAATTTGTGCTTCATGCGATCCATCTGC		2013
OY	1996	AATGATTTTTCACCTAATGTGCCAGGCTTACGAAAGAGAATGT---TGTAACAAAATC		2052
Db	2014	GATGATTTTTTATCCCAATGTGCTCGATCTTAAGAAAGAAAACCTTCTGCAAAATCTT		2073
OY	2053	AAGAAGCATCGTGAGCCGTTCTAATTAAGAAAAGGCCCTTGAGAGATGCTCTCAAAATGAG		2112
Db	2074	AAAAAATATTTGCGAACCATTTCTACAAAAGAAAGTTTATGAAGATGCTTTAAAGTAGAG		2133
OY	2113	CTTCAAGGAAAATTTGACTGAATAATCTTAAATGTGAACCTTCATTTGAAAGATATTTGACA		2172
Db	2134	CTTCAAGGAAAATTTAATGAATATTAATTAATAATGTGAACCTTCATTTGAAAGATATTTGACA		2193
OY	2173	GTACCGGAAAACGTAAATTAATGTGCTGAATCAGTGCCTTAATGCAAGCTTACACCAAGAT		2232
Db	2194	GTATTTGAAGACGTAAATTAATGTGCTCAATCAGCGTTTATGTAAAGATTAATCCGAAAGT		2253
OY	2233	AACTCTGGAAG--AGTAGTAGAGATGCTAAGAAAAGAACTCTGTGAGAAATTTAGTAA		2289
Db	2254	AAAACTAAAAAGGCCGCTAATTAATAATGTATGAAGAAAGCTTTGTCTTAAATTTAGTGAA		2313
OY	2290	GAACTGGAAGACAGTSCAAAGCAATTACCAACAGAAATTAGACAAACCGGACCTGATCTA		2349
Db	2314	GAGGTGGAACGCAATGCCAAAGTATTACCAACAGAAATTACAGAGCTGGAAAAAAGCTTA		2373
OY	2350	AAAAAAGATTATGAACATATGAGGAACCTTAAGAAACGTGCAAGAGGAAGCAATGAACAG		2409
Db	2374	AAAAAAGATGTTAAGACATATGAGGAACCTTAAGGAAGGGCCAAAAAGCAATGAACAG		2433
OY	2410	TCCAGCTCTGTTTGTCACTCATTAAGAAAAAGAAAGTATGATGTCAAAAAGTAAATAGC		2469
Db	2434	TCCAGCTCTGTTTATATCACTTTGTTAAGAAAAAGAAAGTAAATGATGAAATAATATAGC		2493
OY	2470	AAAAACAAGATTAAGAAATGCCGTTTCAACGGAACCTTCAAGATACCAAAAAATATGTAA		2529
Db	2494	AAAAACAAGATTAAGAAATGCTGTTTCAACGGAATTCAGATTAACCAAAAAATATGTGAA		2553
OY	2530	ATACTACGAGGGGAGTTAAGATGTATCCGTAAACAGAAATTGAAGCTTAAAGCATTTGAT		2589



```

Db      2554 ATACTAGAAAGAGATTAGAGGCACTTGTAACAGATCTGAAGCCAGGACTTTGAT 2613
Qy      2590 TTGCACAGAGATATTGTAAGATATGTAGATTGAAAGAAAGATGTAATTAATTGAA 2649
Db      2614 TTGCACAGAGAGATTGTAAGATATGTAGATTGAAAGAAAGATGTAATTAATTGAA 2673
Qy      2650 TCAGATTGAGAAATTAGAGGATTGCAAGACTTGAAGAAATGCAAAAGATTAA 2709
Db      2674 TCGGATTGCGGATTAAAGACGATTGGTTTAAAGAAAGTGTGGAAAGATTGAG 2733
Qy      2710 AAGCTTGTGGCAATTCGAACCTCTGGAGTGAAAGCCGCAAGAAACGTGACGAAAGT 2769
Db      2734 AAGCATGTCAAGATCTGAAGCTCTGGAGTGAAAGTGCATGAATATGTCACAGAAAGC 2793
Qy      2770 ACAACGACAACTAACAAACAACAACGCTGGCCGATCCGAAAGGCAAGGAATTGCAA 2829
Db      2794 ACAACGACGACACACACACACACGCTTACCGATCCGAAAGGCAACGAATTGCAA 2853
Qy      2830 TCCTTACAGACACACACACACATGGGTTACACAGACATGACACACACACGCTTACT 2889
Db      2854 TCCTTACAGACACACACACATGGGTTACACAGACATGACACACACACGCTTACT 2913
Qy      2890 ATCATCATTCATCATCATCAAAATAAATGATGATCAATCAAGGCGGATGCAACCAAC 2949
Db      2914 ATCATCATTCATCATCATCAAAATAAATGATGATCAATCAAGGCGGATGCAACCAAC 2973
Qy      2950 AAGCTGACGACGAGAGAGAAATGATGATGACAGACGCGAAACGAGTGAAGGCTGAG 3009
Db      2974 AAGCTGACGACAGG-----GATGACGACGAGACGCGAAACGAGTGAAGGCTGAG 3027
Qy      3010 ATGAGTGGTGAATGTCATGATGAGGCGGCTGATAGTACGATGTTATTTCTTCATGAT 3069
Db      3028 ATGAGTGGTGAATGTCATGATGAGGCGGCTGATAGTACGATGTTATTTCTTCATGAT 3087
Qy      3070 TAG 3072
Db      3088 TAG 3090

RESULT 5
AAZ94067
ID      AAZ94067 standard; DNA; 3081 BP.
XX      AAZ94067;
AC      XX
XX      15-SEP-2003 (revised)
DT      05-JUN-2000 (first entry)
XX      XX
XX      Pneumocystis carinii major surface glycoprotein gene HMSG32.
DE      Major surface glycoprotein; MSG; HMSG32; human; pneumonia; diagnosis; ds.
XX      OS
XX      Pneumocystis carinii; sp. f. hominis.
XX      FH
XX      Key CDS
XX      1. 3030
XX      Location/Qualifiers
XX      FT CDS
XX      /tag=
XX      /product= "HMSG32"
XX      /note= "a nucleic acid comprising residues 2836-3081 of
XX      this sequence is specifically claimed in Claim 28"
XX      WO200009760-A1.
XX      PD
XX      24-FEB-2000.
XX      PP
XX      17-AUG-1999; 99WO-US018750.
XX      PR
XX      17-AUG-1998; 98US-0096805P.
XX      PA
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX      PI
XX      Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;

```

```

DR      WPI; 2000-206025/19.
DR      P-PSDB; AAY79169.
XX      XX
PT      Detection of the presence of Pneumocystis carinii in specimens by
PT      identification of major surface glycoprotein (MSG) gene sequences using
PT      two or more oligonucleotide primers derived from human P. carinii MSG
PT      protein encoding sequence.
XX      PS
XX      Claim 27; Page 80-84; 110pp; English.
CC      CC
CC      The present sequence is that of the novel Pneumocystis carinii sp. f.
CC      hominis gene, HMSG32, which encodes a major surface glycoprotein (MSG,
CC      see AAY79169). The gene was isolated by PCR amplification of DNA taken
CC      from an autopsy lung sample of an HIV-infected patient with P. carinii
CC      pneumonia. It is 1 of 7 novel, claimed genes (see AAZ79063-69) of the
CC      invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs
CC      include a highly conserved C-terminal region of approximately 100 amino
CC      acids; this region (also claimed) corresponds to residues 2836-3081 of
CC      the present sequence. Direct detection or amplification of human-P.
CC      carinii MSG-encoding genes, especially by PCR using primers directed at
CC      the conserved region of the genes, provides a sensitive and specific
CC      technique for the detection of P. carinii, and the diagnosis of P.
CC      carinii pneumonia, especially in biological specimens (e.g. blood,
CC      sputum) from immunocompromised patients such as those with HIV infection.
CC      (Updated from 15-SEP-2003 to standardise OS field)
XX      XX
SQ      Sequence 3081 BP; 1231 A; 431 C; 663 G; 756 T; 0 U; 0 Other;
Query Match      52.9%; Score 1624.8; DB 3; Length 3081;
Best Local Similarity 72.8%; Pred. No. 0;
Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;
Qy      1 ATGCGCGGCGCGGCAAGCGGCGAG-----GCACAGGAGACACAGAAATGACATTGAT 51
Db      1 ATGCGCGGCGGCGGCAAGCGGCGAGCTGCAAAACATCAAGGCGCTAGTGTATGATGCT 60
Qy      52 GAGGACATGTTTATGCTTAAATTTCTAAGGAAATGATGATGATGAGCAAGATGCAAA 111
Db      61 GAAGAAATTCCTTTTGGCTTTATGCA---GAAAGAAATATATATGATATGAAATGCAAA 117
Qy      112 AAAAACTAAAAAATTTGCCAAGATTTGATGAGCAAACTAAATATATGAAACAAATA 171
Db      118 AAAAAATTAGAAAAATTTGTAAGACATTTAAGCATCAAGATTTAAAAACCGAAAAAGTT 177
Qy      172 CACAGAAACTTAAAGTTTTTGGGAATGAGAAAAAGATACAAATGCAAAAGAACTG 231
Db      178 CACAAAAACTTAAAGGTTTTTGGAAAAATAAAAACAGATTTAAAAATGCAAGAACTG 237
Qy      232 AAAGCCATATTTGAGAAAAATGTAATCAATCAATCAAGAAAACTTAAAGAAACATTTAA 291
Db      238 AAAGAAAAACTCACTCAAAATATGATGCAATCAAAAGAAAACTTACGAAGCAATCAA 297
Qy      292 AAAAAAATTCAGATTAAAGGATTAAGATTGCAAGAGAAATGAACAACAATGCTATTT 351
Db      298 AAAAAAATTCAGATTAAAGGATTAAGATTGCAAGAGAAATGAACAACAATGCTATTT 357
Qy      352 TTGAGGAGATGATGTTCAAAAAGAACTTAAAGATGATGCAATCTTGAGAAATTAAGTC 411
Db      358 TTGAGGAGATGATGTTCAAGCGGAACTTAAAGATGATGCAATCTTGAGAAATTAAGTC 417
Qy      412 TATCAAAAGAAACGTGATTAAGTTGCGGAAGAGTTCTTTTAAAGAGACCTTCGTAGCAT 471
Db      418 TATCAAAAGAAAGCTGATTAAGTTGCGGAAGAGCTCTTTTAAAGAGACCTTCGTAGCAT 477
Qy      472 CTTAATGATCAGTCATATGTGAAAAAAAATTAAGAGATTGCTGTCATGGGAGG 531
Db      478 CTTAATCAATGAATCTACATGTGAAGGAAAGCTCAAAAGAGTTTGAATAGATGATCA 537
Qy      532 GAAAGTATGATGTTAACAACCTGTGCTGAACGAGAAAGACATGATTAAGATATTTTA 591
Db      538 GAAAGTATGATGTTAACAACCTGTGCTTTATCAAAAATGACGTGCAAAACATTTGTA 597
Qy      592 ATTGAAAAAGATTAAGAGTGGGTACTCTTAAAAACAGATGTTTCAGACGACCTAG----- 646

```



Db 598 TTAGAAAAACAAAAATGTATCTCTTAAACAGAGATGTTTACGACGACTTGAAG 657  
Qy 647 -GAAGTTTAAAAAGAAACATGTCTTGAATTACTCGAACATGCTATTTTACATTGGA 705  
Db 658 AAGATGAGTTACGAGAAATGTTTACCACGCTTGAAACGATGCTATTTTATAGAGG 717  
Qy 706 AATTGCGGAG-----ACGACGATTAATT 729  
Db 718 AATTGTGAAGATATATCAAAATGTATAATATCATCCGAAGACTGTTATGAATATTGCCA 777  
Qy 730 AAATGTATTTGAATTGGAGGAAAAATGCCAAGAACAAATATTGCTTATATGCAACAGGA 789  
Db 778 GTGTGTGATACATTGGCAGTGAATGTGAAGAAATTAAGATTATTTATACATCCGGGA 837  
Qy 790 CCCGATTTTGATCCAACTAGCCAGGAGCTACATATGAGAGATATATAGGCTGGAGAG 849  
Db 838 TCCGATTTTCATCCACTTAAGCTTAAGCTTACTGTAGCAGAAACATAGGACTGGAAAG 897  
Qy 850 TTTTATGAAGAGTAGAGAGAGATGAGTTTATTTATGGAAGAAATCATCTAAGAGATGCG 909  
Db 898 CTTTATTAAGGCGCGCAGAAAGAGTGTTCATATTGGAAAGCCTCCTGTAAAGATGCA 957  
Qy 910 ACAGTTTGTGGCATTTGTGATCCAAAGTTCTAGTCTTAAAAAAGAGACAAAG 969  
Db 958 ACTGCTCTACTGGCCGCTTTGATTCAAATCTAAGTCTTAAGATCAAGTGGTAAAG-- 1015  
Qy 970 AAATCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCCATGATGACATAGGCTTTAGAA 1029  
Db 1016 -AATCGAAAAAGTCTTAAAGATTAAGTAAAGGTTAAAAAGTCATGAATTTTGGGA 1074  
Qy 1030 AGTTTATGTAGAAAAATGTTTAAATGATGAGAACGAAAAATGTGAAGATTGCA 1089  
Db 1075 GATTTTGTAAATCAAAATGTAGCTGCTCAAATGAATGAATGAAGATGAAGTTAAG 1134  
Qy 1090 AATGATATTAACAAACTTGCAAAATTTTCACTTCAAAAAGTCATTAATATGCTTTT 1149  
Db 1135 AAGGAGTTAGCAAAAGTAACTAAATCTTTTGAAGAAATTAAGAAATTAACACTCTCT 1194  
Qy 1150 GATCCAAACAAAGAAATTAATGAATTTGTTGAGTGAAGGCTTCCAACTTTCTTAGC 1209  
Db 1195 -----GATCCGAGAAAGTCAATTCATGATTAAGTTTACGACATTTCTTAGT 1242  
Qy 1210 AACGAAGATTGTGCAAAATTTGAGTCTTATTTGTTTCTATTTTGAAGAAAAATGTCAGAT 1269  
Db 1243 GACATGACTGCAAAAGTTTAAAGTCAAGTCTGTTTATTTT---AAAAATGCAACACT 1299  
Qy 1270 GGAGAAATGCAATGTAATAATTAAGACACATGTTACAAAGAGAGCTTGATCAGCG 1329  
Db 1300 CTTTGCAAGAAATGTAATATCTGAAGGACATGTTATAGAGAGGCTTGAAACACAA 1359  
Qy 1330 GCAATTAAGTCTGCAAGAAATATGCGAGGAATGTTACATGTTTCAAAACAAAGCTGG 1389  
Db 1360 GCTAATGAAGCATTTGCAAGAAAAAGATGTACGACGCTTCTATGCTTACAGGCAAAAGATG 1419  
Qy 1390 CTTGAAAACTTCAACAGAAATTAAGTAATGATGAGAACTGAAAAAGAAAAA 1449  
Db 1420 TTTAAGAACTACTGAAAAAATAATGAAAGATGTCGAACCTTAAAAA----- 1470  
Qy 1450 GGAAGTTTCTCAAGAGATTAATTTATTTCTGTGTGTAACGCCAGCAAAAGCACCCGG 1509  
Db 1471 -----ACAAAGGATGAGTTGTTTGTATGTAATGATCCCTTAAAGCACTGCA 1521  
Qy 1510 TTGCTTACATGATCTTGAATGAAAACTATCTTTTACGACAACTGATCAAAAG 1569  
Db 1522 ATACTTTCAGTGAATTCACAGCAAGAGAGTCTTTTTCGGAACAAATTTGATCAAAAG 1581  
Qy 1570 CGAGATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCAGATTTAGGA 1629  
Db 1582 CGAGACTTTCCAAAGACAAAGATTTGCAAGAAATTTAGAAAGAAAGTGAACCTTTAGGG 1641  
Qy 1630 GAGGATTCAAAAAAATTAATGAGGCAATGCTATACACTGAGAGCAATGCAATGCTTTG 1689

Db 1642 AAGATTCAAATCAAGATTAAGTGGCCATGTCTATCGCTTAAAAACAAGTGTGATCGCTTG 1701  
Qy 1690 GGAAGTACAGAAATTTTAAAGCAGGTTTATTTGATGAACACAAAGATACTTTGAAGAC 1749  
Db 1702 GGAAGTACAGAAATCTTGAAGCAGGTTTACTAATGATGAACAGAGATACTTTAAGAACT 1761  
Qy 1750 CAAGAAAGTTGTGTAATAATTAACCTTAAAGAAAAAGTGTATTAATGCTCTAGAAAGAGAT 1809  
Db 1762 CATGAAAACTGTAGAAATATTTAAAGAAAGAAATGTATTAATGCTCTAGAAAGGCTGAT 1821  
Qy 1810 GAGGTTTCTCTTTGATGTGTCTTCAAAAGCCTACGTTGAGCTGATGTTAAAGAC 1869  
Db 1822 GATGTTTCTCTTTGATGTGTGTACCAAAAGCCTACGTTAAGCTGATGTGATGAT 1881  
Qy 1870 GTGAAGACAGTGTGAAGTATTCAAAAAATAATTAAGGCTTATATTTAAATTT 1929  
Db 1882 GTGAAGACAGTGTGAAGTATTTGAAGAAAAATGCAAGCTCAGATATTAATTAATCT 1941  
Qy 1930 CTTGAAAAATTAACAAATTAATAACAACACTGAAAGAAATTTGCTCTTGGCATACG 1989  
Db 1942 CTTAAAAATTAACAAATTAATAACAAGATCAAGCAAAATATTTGTCCCTCATGCAACCA 2001  
Qy 1990 TATTGCAATGATTTTCACTTAATTTGTCCAGGCTCTTACGAAAGAAATAGT---TGTACA 2046  
Db 2002 TACTGCGATGATTTTACCCAAATTTGCTGATCTTAAAGAAAGAAAACTTTCTGTCAA 2061  
Qy 2047 AAAATCAAGAAAGATCGTAGCCGCTTCTATTAAGAAAGAGCCTTGAAGATGCTCTCAA 2106  
Db 2062 AATCTTAAAAAATTTGGAACCAATTCACAAAAGAAAGTTTAAAGATGCTCTTAAA 2121  
Qy 2107 GTAGAGCTTCAAGAAATTTGAATTAATCTTAAATGTGAACCTGCTTGAAGAAAT 2166  
Db 2122 GTAGAGCTTCAAGAAATTTGAATTAAGTAATGAATTAATGTGAATGTGCTTTGAAAGAT 2181  
Qy 2167 TGTACAGTAGGAGGAAAGTAAATTAATGCGTCAATCAAGTGTGCTTATGCAAGCTAACAC 2226  
Db 2182 TGCACATATTTGAAGAAATGTAAGTATTCATCAATCAACAGTTTATGTAAGATTAATCC 2241  
Qy 2227 AAGATTAATCTGAAAGAGTATGAGATGCTTAAGAAAGAACTGTGAGAAATTAATG 2286  
Db 2242 GAAAGTAAATCTAAAAAGACGATTAAGAGTTGAAGAAAGCTTTGCTAAAAATTAATG 2301  
Qy 2287 AAAGAGTGAAGAACAGTGAAGCATTTACCAACAAATTAAGCAACCGGACGCTAT 2346  
Db 2302 GAAGAGTGAAGAACAGATGTAATGTTACACAGAAATTTGAGCAAGAGAAAGAAAC 2361  
Qy 2347 CTAAGAAAAAGTTAAGACATATGAGAACTTAAGAAAGCTGACAGAGAAAGCAATGAAC 2406  
Db 2362 CTTAAAGATGATTTGAACATTTGAAGAACTTAAAGAAAGCGACAGAAACAAATGAAT 2421  
Qy 2407 AAGTCCAGTCTTGTGTGTCATCTCATTAAGAAAAAGAAAGTAATGATCAAAAAATGAT 2466  
Db 2422 AAATCCATCTTGTTTTATCATTTGTTAAGAAAGTGAATAATATCATCGAAAAATGAT 2481  
Qy 2467 AGCAAAACAAAGATTAAGATGCGGTTTCAACGGACTTCAAGATACCAACAAACATGTG 2526  
Db 2482 AGCAAGACAAAGATTAAGAAATCCGTTCAACGGACTTCAAGATACCAACAAACATGTG 2541  
Qy 2527 AAAATACAGAGAGAGGATTAAGAGTATCCGTAACAGAAATTAAGAGCTTAAGCACTT 2586  
Db 2542 AAAATACAGAGAGAGGATTAAGAGTATCCGTAACAGAAATTAAGAGCTTAAGCACTT 2601  
Qy 2587 GATTTCAGCAGAGATTAATTTGGAAGATGATGATTTGAAGAAAGATGTAATTAATG 2646  
Db 2602 GATTTCAGCAGAGATTAATTTGGAAGATGATGATTTGAAGAAAGATGTAATTAATG 2661  
Qy 2647 GAATCAGATTGCAAGATTAAGAGATTTGCAAGATTTGAAGAAAGTATGCAAAAAAGTT 2706  
Db 2662 GAATCAGATTGCAAGATTAAGAGATTTGCAAGATTTGAAGAAAGTATGTCGAAAGATA 2721  
Qy 2707 AATAAGCTTGTCCCAATGGAAGCTCTGAGAGTGAAGCCGCAAGAAACAGTGAACA 2766  
Db 2722 CAAGAGATGTTGAAATTTAAACCACTGAAGTGAAGCCGCAAGAAACAGTGAACA 2781

```
OY 2767 AGTACAAGCACTATCAACAACAACACCGTGGCCGATCCGAAAGCAACGGAATGC 2826
DB 2782 AGCAACAAGCAAGCAAGCAACAACGACCTTACTGATCCGAAAGCAACGAATGC 2841
OY 2827 AATCTCTTACAGACAGACACATGGGTTTACACAGACATGACACACACAGACGCTT 2886
DB 2842 AATCTCTTACAGACAGACATGATTCATGATTCACAGACTTGACACATACACAGACGCTT 2901
OY 2887 ACTATCATCTTACATCACTCAAAATAAATTGACATCAACGAGCCGATGCAACCA 2946
DB 2902 ACCATCATCTTACATCACTCAAAATAAATACTCATCAACAGCCGTTGCAACCA 2961
OY 2947 ACCAAGTGTACGACAGAGAGAGATGATGACAGAGACGTAACCGAGTGAAGGCGCTG 3006
DB 2962 ACCAAGTGTACGACAGAGAGAGATGATGACAGAGACGTAACCGAGTGAAGGCGCTG 3015
OY 3007 AGCATGATGGGTGGAATGTGATGAGGGGGGTGATAGTGAATGTTATTTCTTCATG 3066
DB 3016 AAGATGATGGGTGGAATGTGATGAGGGGGGTGATAGTGAATGTTATTTCTTCATG 3075
OY 3067 ATTTAG 3072
DB 3076 ATTTAG 3081

RESULT 6
AA294063 ID AA294063 standard; DNA; 3042 BP.
AC AA294063;
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
DE Pneumocystis carinii major surface glycoprotein gene HMSGp1.
XX
XX Major surface glycoprotein; MSG; HMSGp1, human; pneumonia; diagnosis; de.
XX
XX Pneumocystis carinii; sp. f. hominis.
XX
XX Key Location/Qualifiers
XX FH 1.3042
XX CDS /*tag= a
XX /product= "HMSGp1"
XX /note= "a nucleic acid comprising residues 2894-3042 of
XX this sequence is specifically claimed in Claim 28"
XX
XX WO200009760-A1.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99WO-US018750.
XX
XX 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
XX
XX WPI; 2000-206025/18.
XX
XX P-PSDB; AAY79165.
XX
XX Detection of the presence of Pneumocystis carinii in specimens by
XX identification of major surface glycoprotein (MSG) gene sequences using
XX two or more oligonucleotide primers derived from human P. carinii MSG
XX protein encoding sequence.
XX
XX Claim 27, Page 49-53; 110pp; English.
XX
XX The present sequence is that of the novel Pneumocystis carinii sp. f.
XX hominis gene, HMSGp1, which encodes a major surface glycoprotein (MSG,
XX see AAY79165). The gene is 1 of 7 novel, claimed genes (see AA279063-69)
```

```
CC of the invention that encode human-P. carinii MSGs (see AAY79165-71). The
CC MSGs include a highly conserved C-terminal region of approximately 100
CC amino acids; this region (also claimed) corresponds to residues 2894-3042
CC of the present sequence. Direct detection or amplification of human-P.
CC carinii MSG-encoding genes, especially by PCR using primers directed at
CC the conserved region of the genes, provides a sensitive and specific
CC technique for the detection of P. carinii, and the diagnosis of P.
CC carinii pneumonia, especially in biological specimens (e.g. blood,
CC sputum) from immunocompromised patients such as those with HIV infection.
CC (Updated on 15-SEP-2003 to standardise OS field)
XX
XX Sequence 3042 BP; 1307 A; 381 C; 606 G; 748 T; 0 U; 0 Other;
XX
XX Query Match 35.8%; Score 1098.6; DB 3; Length 3042;
XX Best Local Similarity 63.0%; Pred. No. 2,7e-217;
XX Matches 1954; Conservative 0; Mismatches 1054; Indels 93; Gaps 13;
OY 2 TGGCGCGGCGGCTCAAGCGGCGAGCGGACACAGAAATAGCATGTATGAGAA---C 58
DB 2 TGGCGCGGCGGCTTAAGCGGCGAGGTAACAGAGCATCAGAGTAGATGAGAGGAAATGC 61
OY 59 ATGTTTACGTTTAATCTAAAGGAAGATGACCTAAGTGAAGCAATGCAAAAAAATC 118
DB 62 GTCTTTGGCTTAAATCTAAAGGAAGATCTAAGATGATTAATAATGCGAAGAAAAT 121
OY 119 TAAATAAATATGCGCAAGATGCTGAGCAAACTAATATAGAACAAATACACAGAA 178
DB 122 TAGAAAAACATTCGAAGAAATTAAGTAGACAAATCTAATCCGAAACAGTACATGAAA 181
OY 179 AACTTAAAGGTTTGGCAAGATGAAAGACATACAAATGCAAAAGAACTGAAGCCCA 238
DB 182 AGTTAAAGATTTCTGTATAGCAAAAAACGTGATTAATAATGTAAGAACTAATAAAAAA 241
OY 239 ATATTGAAAAAATATGATCTAATCAAGAAACTTAAGAAGCAATTAATAAAAAA 298
DB 242 ATGTTGAAAAAATATGCGGATTTTAAACAGATTAAGAAATGCTGTAAGAAAGAG 301
OY 299 TTCAGATTTATACGGAATGATGATTCGAAGAAGATGAACAAATGCTATTTTGGAGG 358
DB 302 CTTCAAAATTTGAAAAATGATGAGCTGACMAAAATGAAACAACTGCTTTTGAAG 361
OY 359 GAGTATGTTCAAAAGAACTTAAGATGATTCGAATCTTGAAGAAATGATGATCAAA 418
DB 362 AAGCATGCTC---TGATCTTACAAAGAAATGCAACGATTAAGAACAATAATGATCAGA 418
OY 419 AGAAAGTGAATTAAGTTGCGGAAGAACTTTCTTAAGAGCACT---TCGTAGCATCTTA 475
DB 419 ATAGCGGTGATTAAGTGAAGAAAGAACTTTTAAAGATTAATAAGAAAGAAATTTTA 478
OY 476 ATGATCAGTCATGATGAAAAAATACTTAAGAGATTTGCGCTGTCATGGGAGGAGAA 535
DB 479 AAGATTAATAATTCATGTAATACTGAAAGATTAATGTAAGAAATTAAGTCAATGA 538
OY 536 GTGATGATTAACAACCTGTGTCTGAACGAAAGAGACATGTAAGATTTTAAATG 595
DB 539 GTGACGATGATGAATTAATGTTGATCAAAAAAATACGTGTGATTAATCTTGTAAAG 598
OY 596 AAAAAATTAAGATG-----CGTACTCTTAATAAGATGTTTCAAGACATGAG 646
DB 599 AAGCGCAACAAAGTGAATCTTTCAAAATCTTAAGAGGAAATTAACAATTAAGG 658
OY 647 GAAGTTTAAAGAAAGAACTGCTGAATTAATGCAATGCTATTTTACATGGA 706
DB 659 AAGATGAACACTTAATAAATAATGCGCATTTATATGAAATGCAATTTTATGATG 718
OY 707 ATTGCGAGAGCAGATATATTAATGATATGATGAGAGAAATGCGCAAGACAA 766
DB 719 AAGTTGTGGAAGAGATCACTGAAGTGAATGGAATAAATAATGTCACAGAGAAA 778
OY 767 ATATTGCTATATGCAACGAGCCGATTTGATCCAACTAGGCGAGAGCTCAATAG 826
DB 779 ATATTACTTACACATTTATATATTCAGAGTTTGAATCTTAAGAAACAGAAATTAATG 838
```

Oy	827	TAGAGGATATATAGGCTGGAAAGTTTTATTAAGAGGTAGAGAGATGGATTTTATATG	886
Db	839	CAGAAAGATGTGACTTAGAAGGAATTTATTAAGAAAGCGCAAGAAAGAACTCTTGTTG	898
Oy	887	GAAAGAAATCATCTTAAGAGATCGCAGCTTTGTTGGCACTTGTGTATCCAAATTTCTAGTC	946
Db	899	GGAAACCTTTACAGCAATGCTACTGCTTTGGTGGCACTTTTGTATTCAGATTCATCTC	958
Oy	947	TTAAAA--AAAGAAGCAAGAGAAATCGAAGAAAGCCCTTCAAAAAAGCTGCAAA	1003
Db	959	TTACAACCTCAACGAATCAACAAAGAAAATGTAAAAAAATCTTGAAGATTAAGTAAAA	1018
Oy	1004	ATCCTCATGAACATGAGGCTTTAGAAAGTTTATGTAAAGAAAAATGTTATGTAATGATG	1053
Db	1019	ATTTAAAAAGAACATGATATATTAATAAAGGCTATGAGAGATTTATTAATGCAAAATAAGATA	1078
Oy	1064	GAAACGAAAAATGTGAAAGATTTGCAAAATGATATTAACAAACTGCAAAATTTTCACCT	1123
Db	1079	AGGACAAAAATTTGCGAAGAACTTAGTAACAGATTTGAAGAAACATGTBAAATTTTTCATTT	1138
Oy	1124	CAAAAGTCATTAATATATGCTCTTTTGATCCAAACAAAAGAAATATGAATGTGTGAT	1183
Db	1139	CAAAAACCTTATGATGATCATTTTGTGGCATGGAATPAAAATGATGGAATATTAATAT	1198
Oy	1184	GGGAAGGTTTGCCCAACATTTCTTAGCAACGAAGATTGTGCGAAATTTGAGTCTATTGTT	1243
Db	1199	GGGGGAATTTATCAACGTTTCTTAAGCAATTAAGATTTGACAAAAATTAAGATCGATTGTCTC	1258
Oy	1244	TCATATTTGAAAAAAAATGTCAGATGAGAAAAATGCATGTAAAAATTAAGACCAACAT	1303
Db	1259	TTTATTTTGAAAAAAGCTGTAGA--AGCGAAATCTGATGCAAGATTTGAGAGACGAT	1315
Oy	1304	GTTAACAAAAGAGACTTGATGATCAGCGGCAATTAAGTGCCTGCAAGAAATATGCGAGAA	1363
Db	1316	GCTACAAAGAGAGACTTGACACATTAAGCAATGAAGTATTAACAAAAGAAATGCGAGAA	1375
Oy	1364	TGTTACATGTTTCAACAAAGAGCTGGCTTGAAGTTTCAACACAGATTTAGTAAAAAGAT	1423
Db	1376	TGCTGCAATGTTCAAAATTAACATGGCTTAGTGTTTCCAAAAAAATCTCATAGAAAGTGT	1435
Oy	1424	GTGAGAAATCTGAAAAAAGAAAAACAAAGAGTTCCTCAAAAGATGAATTAATTTATCTGT	1483
Db	1436	GCAAAAAAGTGAAAAAAGAAATTAAGAGATTTTCCAGTAAATGAATTAATTTGTCTTAT	1495
Oy	1484	GTGTACAGCCAGCAAAAGCAGCCCGGTTGCTTACATCATGATCTTTCGATGAAAACTATCT	1543
Db	1496	GTGTACAACTCAAAAGCAGCTGATTTGCTTTGCGATGATCTTCGATGAAAACTATCT	1555
Oy	1544	TTTTACGACAACTGATCAAAAGCGAGATTTCCGACAGATBAAATTTGCAAGAAAT	1603
Db	1556	TTTTGCAAGACGATTTTGAACAGAAAGCCAGAGTTTTCCAGTBAAGAAAGACATGCAAGAAAT	1615
Oy	1604	TGGGAGAAAGTGGCCAAATTTTAGAGAGATTCGAAAGAAATTAACATGGCGATGCTATA	1663
Db	1616	TATTTAAAGAAATGTGAGGCTCTAGAAAGGATTTCAAAAAATTTGAATGGCGCATGTCTATA	1675
Oy	1664	CACGTGAGCAGCAATGCAATGCTTTGGGAGCTACAGAAATTTTAAAGCAGGCTTTTATTGG	1723
Db	1676	CATTAGCGCAAAATTTGTGATCAATTTGABAAACGCTBAAGAAATTTGAAAGAACTTTTATCTAA	1735
Oy	1724	ATGAACACAAAGATCTTTGAAGACCAAGAAAGTTGTGTAAATAATCTTAAGAAAAAGT	1783
Db	1736	ATGAACATTAAGGATATATTTGAAAAATCAAGAGAAATTTGTGAATGTATTTGAAGAGAAAT	1795
Oy	1784	GTAATTAATGTCTTAGAAGAGAGATGACGTTTCTCTTTGTATGTGTCTTCCAAAACG	1843
Db	1796	GCAATGAATGTCTTAGAAGAGAGAAATGAACGTTTCTCTTTTATGTCTTTGCAAAATA	1855
Oy	1844	CTACCTGTAGCTGATGTTAAAGACGTGAAGACAGGCTGTGAAGTATTTCAAAAAATA	1903
Db	1856	GGACTTGCAGAAATATGTGTAGAAAGTGTGAAAAATCAATGCAAAATATTTGAAAAAACA	1915
Oy	1904	TAAAACTTCAATATTAATTTGAATTTCTTGAATAATATACAAATTAATAACAACCTGG	1963

Db	1916	TTAATAAAATACCAAGTAATG-----ATAGTAAATCTAAATATGAAAGATTAG	1963
OY	1964	AAAGAAATGTCCCTCTGGCATACGTAATTGCATATAGATTTCACTTAATGTCCAGTC	2023
Db	1964	GGACATATTGTCTAATTGGGACCCACACGTCCTATGATTGGACCAATGGCCGGATC	2023
OY	2024	TTACGAAAGAGATATGTTGTACAAAATTCAAAGAACCTGTGAGCCGTTCTATTAAGAA	2083
Db	2024	TTGAAAAA---AAATAATGTGAACCTTGAAAAATTTTCCAAACCTTATTATTAAGCAA	2080
OY	2084	AGGCCTTGGAAGATGCTCTCAAGTAGAGCTTCAAGAAAAATTGACGATTAATCTAAAT	2143
Db	2081	GAGACTTGAAAAATGCACTTATATTTAGATTTAGAGACATCTTGATTAAGAAAAAACT	2140
OY	2144	GTGAACCTGCATTGAAAAAGATATTTGTACAGTAGCGGAAAAGTAAATATGCGTCAATCA	2203
Db	2141	GCAAAAAATCTTGATTAAGTACTGTACATATGGATTCAACAGAAATTAACAACCTTA	2200
OY	2204	GTGCTTATGTCAAAAGCTTAACCAAGGATTACTGTGAAAAGATGAGAGTCTTGAA	2263
Db	2201	AAAGTTTTTGTAT-----ACAGTTCTACGTATACAAAGAAACATTTTGAG	2245
OY	2264	AGGAACCTGTGAGAAATTAAGTGAAGAAGTGAAGAACGTGCAAGCATTTACCAAG	2323
Db	2246	ATTAACCTTGGAAAACTAGTTCAGCGGTGAAAAGAAAATGCAAGATTTCAAAAG	2305
OY	2324	AATTAGCAACACCGCAGCTGATCTAATAAAAAAGTTATTAAGACATATAGAGAACTTAAGA	2383
Db	2306	AACTTGA AAAAGCAAAAATGATTTAAGAAAAACATTAAGATTATGAAAAAGTAAAAA	2365
OY	2384	AACGTGCAGAGAAAGCAATGAACAATGCCAGTCTGTTTGTCTCATATTAAGAAAAACG	2443
Db	2366	AGGATTA CA AAAAATGCAATGGAAGAAAA CAATCTCGTTTTTCAACAACCTAAATCAACAG	2425
OY	2444	AAAGTAATGTATCAAAAAGTAAATGCAAAAAACAAGATTAAGAACGCGTTTCAACGGAAC	2503
Db	2426	ATAATTA AAAACGAAAAAGG-----AGTCAAGCCTTAGTACCGCTTAGTGTAG	2470
OY	2504	TTCAAGATTAACCAAAAAACATGTGAAAAATACTACGGAAGGGAGTTAAAGATGTATCCGTAA	2563
Db	2471	TTCAAGATTAATGTATCAATTT-----AAACTGTGAAAAAGAAATGAAAAAGTTCAATGA	2544
OY	2564	CAGAAATTAAGAGCTAAAGCATTTGATTTGGCAGCAGAAATTAATTGGAAGATATGTAGATT	2623
Db	2535	CAGAAAAAGAAAGCAAAAGCGTTGATTTGGTAGCACTAGCATTCAGCTTTATTTAGAGT	2584
OY	2624	TGAAGAAAGATGTAAATTAATTGSAATCAATTTGCAGAAATTTAAGAGAGATTGCAAAAGACT	2683
Db	2585	TAAAGAAACGTGTACATCTTAAGAGATGATTTGCGAATTTGAAAAAGAAATGTAAA--T	2641
OY	2684	TAGAAGAAATGTGAAAAAGATTAATTAAGCGTTGTCCCAATCTGAACGCTCTGGAGGTGA	2743
Db	2642	GTAAGACCAAGTGC AAAAGAGATTAAGAAAAATATGTTTAAAAATTAAGAACCACTGAAGATTA	2701
OY	2744	AGCCGCAAGAAACAGTGA CAGAAAAGTACAACGACAATCAACAACAACAACAACGCTTG	2803
Db	2702	AGCCACATGAAATTA AAAACAGTAAACGAAACCAACATTAACAACAGTCA CAGAAACAGTCA	2761
OY	2804	CCGATCCGAAGGCAACGG-----AATGCAATCTTTACAGACACACAGCA	2848
Db	2762	AAAGAAAGCAAAAAACAGTAGAGAGACGGAAGAAATCCAATCTCTAGCAACAACAGACA	2821
OY	2849	CATGGGTTTACACAGACATCGACACACACAAGACAGTGTACATACATCTTACCATCAAT	2908
Db	2822	CTGGGTTCACAAAACCTCAACCAATTCACAGCACTTCCAGCTACCTCCAGATTACGT	2881
OY	2909	CAAAAATTAACATTGACATCAACGAGCGATGCAAAACCAACCAAGTGTACGACAGAGAGCG	2968
Db	2882	CAAGATTAACCTACCTCGACGAGGGGGGTGAAGCTTACGAAGGTATACGACAGAGAGGG	2941
OY	2969	AAAGTATGTCAGAGACGTGA AACCGAGTGAAGGGCTGAGAGATGATGGGTGAATGTGA	3028



1133 CTAATATATGCTTTTGTGATCCAAAGAAATGAAATGTTGGATGGAGGCT 1192  
1124 GGATTTGAAATATTTCCCTCCAGCTAATGTAAA---GGAAATATTTGGTCATATGT 1180  
1193 TGGCAACAATTTCTTTCGAAACGAAGTGTGGAAATTTGGAGTCCATTTGTTCTAATTTG 1252  
1181 TGCATACATTTCTTGTGTAAAGAGAGTGTACGAAACCTGTGTGGATGTTTTTAATTTGA 1240  
1253 AAAAAAATGTCAGATGAGAAAAATGATGTAAAAATATTAAGACAATGTTACAAA 1312  
1241 AAAGCCAAAGCTCACTT---GAAAGCCCTGCAATTAACCTTAAGACACATGTTATAAA 1297  
1313 GAGACTTGTATGACGCGCAATTAAGTCTGCAAGAAAATATGCGAAGATGTTACATG 1372  
1298 AAGGCTTGAAGCAGTACCAATGAAAGATTACAGATAGATTAGCGGAAAAATTTGCAAG 1357  
1373 GTTCAAAACAAAAGCTGTGAAAAGTTTCAACAGATTTAGTAAATGATGTGAAAC 1432  
1358 GTTCAAAATGAAACATGCGCTTGAAACCTTCAAAAAACCTGTGAAAAGTTGTGAAAAA 1417  
1433 TGAATAAAGAAAAAAGAAAGTTTCTCAACGATGATTAATTTATCTGTGTACAGC 1492  
1418 CGAAAGAGA-----AAGTGAATTAATTTGTACTATGTATGAACC 1459  
1493 CAGCAAAAGACGCCGCTTCTTCAACATGATCTTGAATGAAAATATCTTTTTCAGAC 1552  
1460 CAATTAATAACGCGCTTACAGTGTCAACAGATTTGCAATGAGGCAAGTTCCTTTGCAAG 1519  
1553 AAACAATGATCAAAAAGGAGATTTCCGCAAGATTAATAATTCAGAAATTTGGGAGAAA 1612  
1520 AGCATTTAAGCAAAAAAGAGATTTTCCAAACAGAAAGAGATTGTAAGAAATTAAGAAAA 1579  
1613 AGTCCAAAGATTTAGAGAGATTCAAAAGAAATTAATCATGCGCATGTACACTGAGC 1672  
1580 AATGTAGAGTCTTATGAGAAAAGTTCAAGAGAAATTAATGTGCAATGTATACGTTAAAA 1639  
1673 AGCAATGCAATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACA 1732  
1640 AGCATTTGCAATCGCTGAAAGCATAGAGCACTTGAAGAGAGATGCTTAAAAAGAAATA 1699  
1733 AAGATACCTTTGAAGAACCAAGAAAGTTGTAAATTAACCTAATAAAGAAAGTTAAT 1792  
1700 AAGGATATTTTAAAGATGAAATAGCTGCAAAAGAAAGCTAAGAAACGATGTGAAAAAT 1759  
1793 GGTCTAGAAAGAGATGACCGTTTCTTTTGTATGTGTCTTCCAAAACGCTACGTGTG 1852  
1760 GGTTTAGAAAGAAAATAAATTTTTCGCTTGTGTGACTTGGAACTTGTGTGCA 1819  
1853 AGCTGATGATTAAGACGTGAAAGACAGGTGTGAAGTATTCAAAAAAATATTAAGCTT 1912  
1820 AAAAGATCACTGAAATGTGAAATCTAAATGTAATATATGAAAGACATATGAAACTA 1879  
1913 CATATATTAATGAATTTCTGAAAATTAATACAAATTAATTAACAACCTGAAAAAATTT 1972  
1880 TGAACGTTATTAAGTAAGTAAGTAAAAAGAGAAAAAATATGTAATTT-----TT 1930  
1973 GTCCCTCTTGCGCATGATATGCAATGATTTTCACTAATTTGTCAGAGTCTTACGAAG 2032  
1931 GGGCTCCATATTTGTAAAAAGTACGAGCAAAATTTGAAAAAATCTTAAAAACGAGAAAAAG 1990  
2033 AGAATAGTTGTACAAAAATCAAGAAAGCATCTGAGCCGTTCTATATAAGAAAGGCTTTGG 2092  
1991 ATGGCAATGCAAAAAACCTCAATATAAAAGTCAAAATCAATCTTGAATAAGAAAGCTTTAG 2050  
2093 AAGATGCTCTCAAAAGTACAGCTTCAAGAAAAATGACATGATTAATCTAATGTGAACCTG 2152  
2051 AAAAAATTAAGTTGTAGAAAGATTTGAAGGATGTTATCAAAACGTAGAGAAATGTAACAATA 2110  
2153 CATTTAAAGATATTTGTACAGTAGCGGAAAGCTAATATGCTCAATCAGTGGCTTAT 2212  
2111 CACTTAATATATCTGTACACATTTGAAAAAGGCAAGAAAGTGGTTGAAACTTTGT--- 2167

2213 GCAAAAGCTTAACACCAAGATTAATCTGTGAAAAGAGTGAATGAGATGCTTGAAGAACTCT 2272  
2168 -----GCAAAAGGAAAAAAAC---CAAGAGTGCATTTAAAGTTAAGAAAGAACTCT 2218  
2273 GTGAAATATTAAGTAAAGAAAGTGAAGAAACAGTSCAAAGCATTAACCAACAGAAATTAAGAC 2332  
2219 GTGAAAAGCTAATTAACGTTAATAAGAAATATCTCAAAATTTAAGAGACAGCTTGAAG 2278  
2333 AACCGGAGCTGATCTAATAAATAAGATTATAGACATATGAGAACTTAAAGAAAGCTGCAAG 2392  
2279 AAGTAAAAAGAGTCTTGAATAAAGAAAGAAAGAAAGTATTAATAAATAATTAAGAAAGAGAG 2338  
2393 AGAAGCAATGAACAATCCAGTCTTGTTTTGTACATCATTTAAGAAAAACAAAGTATG 2452  
2339 AAAAAAGCCTGGAAGATGCAAACTTATTTATGAGAGCGAAAGGAGCTGATATATAT- 2397  
2453 TATCAAAAAGTATATAGCAAAAAAACAAGATTAAGATGCGTTTCAACAGCACTTCAAGATA 2512  
2398 -----AATATATAGTCACTAATTAATAAGATCTCATCTGATACACTTAAGGAAGAAAGGCA 2452  
2513 CCACAAAACATGTGAAAATATCTACGAGGAGGATTAAGAGATATCCGTAAACAGATTAAG 2572  
2453 CAACAGATTTT-----AACTGTAAAGAAATGCAAAAGTGCATGTATACAGAAAAAG 2506  
2573 AAGCTAAAGCATTTGATTTGGCAGCAGAAATTTGCAAGATATGTATGATTTGAAGAA 2632  
2507 AATTAGCAGCATTTGATTTGTGTACCAAGAGCATTTGTCTCTATCTGAATTTGAAGAAA 2566  
2633 GATGTAATTAATTTGGAATCAGATTTGAGAAATTAAGAGATTTGCAAGAACTTGAAGAAAG 2692  
2567 TATGTATATCATCTCAAGAGATTTGTGTTTCAAAAAAGAGT---TGACTGTAGAGATC 2623  
2693 TATGCAAAAAGATTAATTAAGCTTTGTGCAATCTGAAGCCCTTGAAGGTGAAGCCGACG 2752  
2624 CATGTAATAAGATACAGGAAATATGTTCAACATTAAGCCACTAAAGTGAAGACACAGC 2683  
2753 AAACAGTGAAGAAATGACAAAGCAACCACTACAAACA-----ACAAACAACGTTGCCG 2806  
2684 AATATGTAATCTTAAAAACATTAACAATCAACCAACCAACCAACCAACCACTACATTAAG 2743  
2807 ATCCGAAGGCAACGGAATGCAATCTTACAGAACCAACAGACATGAGTTACAGACAT 2866  
2744 ACGAAAGGCAACAGATCTGCCACTTTTACAGAACCAAGATCTGGTCCAAAGAGCT 2803  
2867 CGACACACACAGCAGCTCTAATCAATCTACATCACTCAATCAAAAAATTAACATTTGACAT 2926  
2804 CGACCCATTAAGCACAATCCCAACACATCTACAGTCAAGTAAGTAAGTTGACCT 2863  
2927 CAACGAGCCATGCAAAACCAACCAAGTGTACAGACAGAGAGAAATGATTCAGAGACG 2986  
2864 CGACAAAGCAGGTGTAAAGCTTACGAAGGTGTACAGAGAGAGAAATGAAACAGAGAGCG 2923  
2987 TGAACAGAGTGAAGGCGCTGAGATGAGTGTGAATGTGATGAGGAGGAGTGAATGAG 3046  
2924 TGAAAACGAGTGAAGGCTTGAAGTGAATGATGAGTGAATGAGGAGGAGTGAATGAG 2983  
3047 CAATGTTATTTCTTCAATGAT 3069  
2984 CAATGACGATTTCAATCAATGAT 3006

RESULT 8  
AA041223  
ID AA041223 standard; DNA; 2110 BP.  
XX  
XX AA041223;  
XX AC  
XX 27-AUG-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 02-SEP-1993 (first entry)  
XX  
XX Clone FC3 encoding major surface gp of rat P. carini.  
XX

XX	Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;
XX	HIV; human immunodeficiency virus; diagnostic; PCR; ss.
XX	
OS	Pneumocystis carinii.
XX	
PN	USN7958683-N.
XX	
PD	01-APR-1993.
XX	
XP	09-OCT-1992; 92US-00958683.
XX	
PR	09-OCT-1992; 92US-00958683.
XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX	
PI	Kovacs JA, Angus CW, Powell F, Edman JC;
XX	
DR	WPI; 1993-159487/19.
XX	
PT	Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
XX	copy gene family - used in a vaccine and diagnostic assay for P. carinii
XX	infection in AIDS patients.
XX	
XX	Disclosure; Page 34; 50pp; English.

Sequence 2110 BP; 745 A; 216 C; 480 G; 509 T; 0 U; 160 Other;

Query Match	9.4%	Score 289.6;	DB 2;	Length 210;
Best Local Similarly	52.0%;	Pred. No. 6.1e-50;		
Matches 941; Conservative	0;	Mismatches 789;	Indels 78;	Gaps 10;

Qy	25	GCAGCAGGAGCACACAGATTCGACTTTGATGAGAAACATGTTTTCAGCTTTAATTCCTAAAGGA	84
Db	351	GCACACAGGAGCACACAGATGACCTTTAAGGAGAACCTTTTGCTTTCATTCGGAAGAAAG	410
Qy	85	GATGAGCTAAGTGAAGCAGGAATGCACAAAAAACTTAAAAAATATTCGACAGATTGACT	144
Db	411	GAATTAATGATTAATGAGGATTTAATGCACAAACGAAGCTCAAGAAATATTTGTGAAGATTGAAG	470
Qy	145	GAAGCA-----AACTAATATATAGAACACAGTACACAGAAAACTTTAAAGCTTTTTCGAA	198
Db	471	GAAGCAGATGCTTAAATTCATATGTTAATGATTAAGTTTAAAGAACTTTGTGTGTGTGAT	530
Qy	199	GATGAGAAAGCAGATACAAATATGCACAAAGAACTGAAGCCATATTTGAGAAAAAATGTACT	258
Db	531	GAAGCAAAACGAGATTAATAAATATGCACAAAGACCTGAAAGACAAAGTTGAAGTGAATTAGAA	590
Qy	259	ACAAATCAAAAGGAAAACCTTTAAGAGCAATTTAAAAAATTCAGATTTTAAACGATTAAG	318
Db	591	AAATTTTGATGATGAACCTTCMAAGACGATTGAAAGA-----CATTTAAAGATGAA	638
Qy	319	GATTGCAAGAGAAATGAACAAACATGCTATTTTTGGAGGAGTATGTTCAA---AAGAA	375

Db	639	AATTGTGAAAAACATGAAAGAAAATGTATACCTTTAGAAACAACGGGTTATAGTAAAGAT	638
Qy	376	CTTAAGAATGATTCGCAATACCTTTGAGAAATAAAGTCATACAAAAGAAACGTGATAAGTT	435
Db	699	ATTAAAGAAAGACTGTGTCAAGTTGAGGAAAGAAATGTTACAAAATTGAAGCCGTAAAAAAGCTG	758
Qy	436	CGCGAAGAAAGTCTTTTAAAGACACTTCGACGATCTTAAATGATCATGTATGTGAA	495
Db	759	GCAGAAGAGCTCCTTTTGAAGGGCCCTCGAGAGGGATCTTAAAGATGAAGCTAAATGTAA	818
Qy	496	AAAAAAGCTTAAAGAGATTTGCCCGTATGAGGGAGGAAAAGTGAATGTTAAACAACTTG	555
Db	819	GAAAAAGATGAAAACGTGTTGGCCCAATGTTAAAGCCGAAAGGTGACGAGCTGATGTTTTC	878
Qy	556	TGCTCGAACACGAAAGAGACATGTAAAGATATTTTAAATGAAAAAGATTAAGAAATGCGGT	615
Db	879	TGCCTTGATTCGATGAGACGTGTAAAGCCGTGAAAAAACAATACAGAAAGAAAGTTGCTG	938
Qy	616	ACTCTTAAACAGATGTTTCAGCAGCACTAGAAAGTTTAAAAAAGAAACATGTCTTGAA	675
Db	939	CCTTTAAAGAAAAGCTTAAAGATGAGCGAAAT-----AAAGGAAAAATGTCATGAA	989
Qy	676	TTACTCGAACATGCTATTTTTCATTTGAAATTCGCGAAGACAGCATATAATTAATG	735
Db	990	AGACTTGAAGAAATGCTATTTTAAAGAAAGCCGTACTGA-----AACAAATGT	1040
Qy	736	ATTGAATTTGGAGGAAATATGCAAGAACAAATATTCCTATATATGCCACGAGCCGAT	795
Db	1041	GATAGAGATATGAAGCAATGCAAGAAAAAGAAATGATTCATATTAAGCCGCGGAATCTGAT	1100
Qy	796	TTTGATCCAACTAGGCGCAGAGCTCAATAGCAGAGATATAGGCGTGAAGAGTTTAT	855
Db	1101	TTTAGTCCTGTCAAGCCGAAGGCGCTGTGTGAGAGTATTTGGGTTTGATGATGTGTAT	1160
Qy	856	AAGAAGGTAGAGGAGATGAGTTTATTTTGAAGAAATCATCTAAGATGCGACAGCT	915
Db	1161	AAAAAGCGTGAAGAAAGAAAGAAATTAATTAATGAAATCAGAGATGATCTACCAAGGAAG	1220
Qy	916	TTGTGGCATTTGATTCCAAGATTTCTAAGCTTTAAAAAAGACGACAAAGAAATGCG	975
Db	1221	TCAGTACAAATTTCTGCAAGATCTCTGTACTGTGAGCAGAGATGAGAAATGATGCA	1280
Qy	976	GAAAGAGCCCTTCAAAAAGAGCTGCAAAAATCTCATGAACATGAGGCTTTAGAAAGTTTA	1035
Db	1281	GGAGAAAGAAATGCGGTTAAAGCGTTAGGAAAATGTGAACCTTAAGTATTTGAATACTGAT	1340
Qy	1036	TGTAAAGAAAAATGCTTTAAGTAAATGATGAAACGAAAAATGTGAAGATTCGAAATGAT	1095
Db	1341	TTGATGAGTTATCCAAAGATGCTGATTAAGAAATTAATGCAAAAAAAGCTAGATG--	1398
Qy	1096	ATTAAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTCTTTTGATCCA	1155
Db	1399	-TAAAGAAAGATATACAAAACCTCAAGTTAATCTTTATGTGAAGGCTGTGTCTACGAG	1457
Qy	1156	ACAAAGAGAAATATGAA-----ATTGTGATGAGGAGGGTTCACCAATTTCTTAGC	1209
Db	1458	TTTAAAGAGATTAATAAATCAACATCTTTTATCGTGGGACAGCTTCCACATTAATTTACG	1517
Qy	1210	AACGAAGATTGTGGAATTTGAGAGTCTTATTTGTTTCTAATTTGAAAAAAAATGTCCAGAT	1269
Db	1518	AAGGAGAGTGTGACAGAACTGAGTCCGAATGTTTCTAATTTAGAAAATGCGGTGAAGAT	1577
Qy	1270	GGAGAAAT-----GCATGTAAAAAATAAGCAACATGTTTACAAAAGAGACTGAT	1323
Db	1578	AATAGATTTGTGAAGCGGTCAAAATCTACGATCAGCGTATTAATAAAGGACAAAGC	1637
Qy	1324	GCAGCGGCAATTAAGTGTGCAAGAAATATGCGAGAAATGTTA---CATGTTCAAC	1380
Db	1638	AGGATGTTGAATTAAGTCTTTCAAAAGAAATTAAGGAAAGCTGTGATGTAAAGTTT	1697
Qy	1381	AAAAGCTGGCTTGAAGAGTTTCAACAGAAATTAATAAGTATGTGAGAACTGAAAAAA	1440



```
Db      1698 TATAGCGATCCTTAAGATTGTAAAAAATATGTGTAGAAAAGTACAAACTTAAAAA 1757
Qy      1441 GAAAAAAGAAAGATTCTCAAAACGATGAATATTATTCTGTGTACAGCCAGCAAAA 1500
Db      1758 GATATAA-----GATACCTTCAAAATGTCTTTATCTTAAGAA 1796
Qy      1501 GCAGCCCGGTGTCTTACACATGATCTTGCATGAAAACATATCTTTTACGACAAACTG 1560
Db      1797 CTATGTATGGGCTTCAAAATGATATTTTCTCCAAATCCAAAGATTAAAGTTCGCTTTTA 1856
Qy      1561 GATCAAAAGCGAGATTTCCCGACAGATTAATAATTCAGAAATTTGGGGAGAAAGTCCAA 1620
Db      1857 GATGATCAGAGAGATTTTCATTGTAAGAAAGATTTGCTTGAATTTGGAGAGAAAGTGTAT 1916
Qy      1621 GATTTAGAGAGAGATTCAAAAGAAATTACATGCCATGTCAATACATGAGACGAACATGC 1680
Db      1917 CAATCTTAGTAGATTTCAATTTATTTGAATTTAGAAAAGTATTAACAATTGAAAAGAGCTGT 1976
Qy      1681 AATCGCTTGGGGAATCAACGAAATTTTAAAGCAGTTTATTTGATGAACACAAAGTACT 1740
Db      1977 GAATATTTTTCAGCTTACAGAAAGATTAGAAAAAGTATTTTGAAAAAAGGATGATTCG 2036
Qy      1741 TTGAAGAACAAGAAAGTGTGTAAATATCTAAAGAAAAGTATTAATAGTCTAGA 1800
Db      2037 TTAATGATTCAGAGAAACTGTACAAAGGATTCATGAGAAATGTATTAATTAAG 2096
Qy      1801 AGAGAGAGA 1808
Db      2097 AGAGAGAAA 2104

RESULT 9
AAQ40201
ID      AAQ40201 standard; cDNA to mRNA; 2814 BP.
AC      AAQ40201;
XX      AC
XX      DT 25-MAR-2003 (revised)
XX      DT 09-AUG-1993 (first entry)
XX      DE Sequence of JPB10 encoding the gp16 major surface antigen of rodent
XX      DE Pneumocystis carinii (PC).
XX      KM Pneumocystis infection; diagnosis; AIDS; vaccine; therapy; sb.
XX      OS Rattus rattus.
XX      FH Key location/Qualifiers
XX      FT m16c_feature 251..2814
XX      FT /*tag= b
XX      FT /*note= "Identical to the homologous human Pc DNA"
XX      FT CDS 1617..2813
XX      FT /*tag= a
XX      PN WO9307274-A1.
XX      PD 15-APR-1993.
XX      PE 30-SEP-1992; 92WO-US008328.
XX      PR 30-SEP-1991; 91US-00768166.
XX      PR 18-OCT-1991; 91US-00761034.
XX      PA (GENO ) GEN HOSPITAL CORP.
XX      PI Fishman JA;
XX      DR WPI: 1993-134463/16.
XX      DR P-PSDB; AAR35082.
XX      PT Isolated DNA encoding antigens of mammalian PNEUMOCYSTIS CARINII - used
XX      PT for prevention, treatment and diagnosis of PNEUMOCYSTIS CARINII
XX      PT infection.
```

```
XX      XX
PS      Claim 4; Fig 2; 74pd; English.
XX      CC The PC DNA, antigens and antibodies can be used for early diagnosis of PC
XX      CC infection, partic. in AIDS patients. The antigens and antibodies can also
XX      CC be used in vaccines against PC. The antigens, antibodies and antisense
XX      CC RNA can also be used in the treatment of PC infection. (Updated on 25-MAR
XX      CC -2003 to correct PN field.)
XX      SQ
XX      Sequence 2814 BP; 929 A; 337 C; 657 G; 891 T; 0 U; 0 Other;
XX      Query Match 8.5%; Score 261.6; DB 2; Length 2814;
XX      Best Local Similarity 51.3%; Pred. No. 4.1e-44;
XX      Matches 879; Conservative 0; Mismatches 764; Indels 69; Gaps 9;
Qy      25 GCAGCAGGACACAGAAATAGCATTTGATGAGAACATGTTTACCTTAATCTTAAGAA 84
Db      1152 GTACAGATGACACAGAAATGACATTTAAGAGGAAACACCTTTGGCTTCATTGTGAAGAC 1211
Qy      85 GATGACTAAGTGAACAGAAATGCAAAAAAACTAAAAAATTTGCCAAGATTCAT 144
Db      1212 AAACATGATGATGAATGAATGCAAAAAAGCTCGAGAAATTTGTAAAGATTTAAG 1271
Qy      145 GAAGCAAAACTAAATATAGAAACAAGTACAGAAAACCTTAAAGTTTTCGAAAGATGA 204
Db      1272 AAAGCAGATGAGATTT---CAGTGTGATGAGAAAGTTAAAGACCTTTGTGATGATAA 1328
Qy      205 AAACGAGATCAAAATGCAAAAGCTGAAAGCCAAATTTGAGAAAAATGTACTACAAATC 264
Db      1329 AAACGAGACGAAAATGCAAAAGCTGAAAAAAAAGTTAAAGATGAAATTTGGAACCTTTT 1388
Qy      265 AAAGAAAACTTAAAGAACATTAATAAAAAAATTCAGATTATTAAGGATTAAGATTTCG 324
Db      1389 GATACGATCTTGAAGCATTCGT-----AATGCAATAGAAAGATGAAGAGATTTGT 1439
Qy      325 AAAGGATGAAACAACATGCTATTTTGGAGGAGATGTTCAAAAGAACTTAAGAT 384
Db      1440 AAAAAACATGAAAGAAATGTATCTTTTAGAGAAAGACACCAATATGCTTTAAGAG 1499
Qy      385 GATTGCAATCTTTAGAAATTAAGTGTATCAAAAGAAACGTGATTAAGTTGCGGAAGAA 444
Db      1500 AACTGTGCAGATTTGAAGGAAGAGATGTTAGCAATTTGAAGCGTTAAAGAGTGCGCAGAGAG 1559
Qy      445 GTTCTTTTAAAGACATCTTCAGCGATCTTAATGATCAATATGTAAGAAAAAACTT 504
Db      1560 CTCTTTTGAAGGGCGCTCGAGGGAGTGTAAAGATGAAGCTTAATGAAGAAAGATG 1619
Qy      505 AAAGGATTTGCCCTGCATGAGGGAGAAAGTATGATTAACAACTGTGTCTGAC 564
Db      1620 AAAAATGTTTCCCAATGTTAAGCCGAGAAAGTGAAGCTGATGTTTCTGCTTGAT 1679
Qy      565 CAGAAAGACATGTAAAGATATTTTAATGAAAGAAAGTGAAGTGCCTACTTAA 624
Db      1680 TCGATGGAACGTGTAAAGCGCTGAAACAAATCAAGAAAGTTTGCCTCTTAATA 1739
Qy      625 ACAGATGTTTACAGACATGGAAGTTTAAAAAAGAAACATGCTTTGAATTTACTGAA 684
Db      1740 GAAAAGCTTAAAGATGCGCAAT-----AAAGAAAAATGTCAGAAAGACTTTAG 1790
Qy      685 CAATGCTATTTTCAATTGAAATTTGGGAGACAGACGATATTAATTAATGATTAATTTG 744
Db      1791 AAATGTCAATTTTACAAAGAAAGCGTGTACTGA-----AACAAATGTGATGAGAT 1841
Qy      745 GGAGAAATGCCCAAGAACAAAATATTTGCTTATATGCAACCGACGAGACCGATTTGATCA 804
Db      1842 ATGAAGCAATGCAAGAAAGAAAGATTCACATATTAAGCGCGGAATGATTTTATGCTCT 1901
Qy      805 ACTAGGCCAGAGCTCAATATGACAGAGATATTAAGGCTGGAAGATTTTATTAAGAGGTA 864
Db      1902 GTCAAGCCGAAGGCGTGTGTTGTAAGATTAATGAGTATGATATTAAGAGGCT 1961
Qy      865 GAGAGATGAGATTTTATTTGAAGAAATCATCTAAGAGATGGACAGCTTTGTGTGCA 924
```



Db 1962 GAAAAAGAGAAATTAATTAATGGAATAATCAGAGATGATCTACCAAGAAAGTACAGTACA 2021  
 Qy 925 TTGTGATCGAAGTCTTACGCTTAATAAAGAGACAGCAAGAGAAATCGAAGAACCC 984  
 Db 2022 AAATTTCTGCAAGATCTTGTCTACGTGTGACAGAGATGAAATGATGACAGGAGAA 2081  
 Qy 985 CTTCAAAAAAGCTGCAAAAAATCCTCATGAAATGAGGCTTTAGAAATTTATGTAAGAAA 1044  
 Db 2082 TGGGTAAAGCGTTAGAAATAATGAAATCTTCAAGATTTGAATACATGATGATGAG 2141  
 Qy 1045 AATGCTTAAGTAATGATGAAACGAAAAATGTAAGAAATTCGAAATTAATTAACAAA 1104  
 Db 2142 TTATGCAAAATGCTGATTAAGAAATAATGCAAAAAAAGCTAGATG---TAAAGAA 2198  
 Qy 1105 ACTGCAAAATTTTCACTTCAAAAGTCACTAATATGTCCTTTTGAATCCAAACAAAAG 1164  
 Db 2199 AGATGTAACAAATCGAAGTTAAATCTTTATGTGAAAGGTTGCTACGAGTTTAAAGAA 2258  
 Qy 1165 AATTAATGA-----ATTGTTGATGGAAGGTTGCAACATTTCTTAGCAAGAGAT 1218  
 Db 2259 GATTAATAATCAACATTTTATCGTGGGACAGCTTCCACATTTATTACGAAGGAGAG 2318  
 Qy 1219 TGTGCAAAATTTGAGTCTTATGTTTCTTATTTTGAATAAATAATGTCAGATGAGAAA-- 1276  
 Db 2319 TGTGCAAAATTTGAGTCTTATGTTTCTTATTTAGAAATGCGTGAATGAATGAGATT 2378  
 Qy 1277 ----ATGCAATTAATAATTAAGCAACATGTTTCAAAAAAGCACTTGAATGACGGGCA 1332  
 Db 2379 GGTGAACCGTGTCAAAATCTACGATCGCGCTATTAATAAGGACAAAGAGATTTG 2438  
 Qy 1333 AATTAAGTGTGCAAGAAATATATGCGAAGATGTT---ACATGTTCAAAACAAAGCTGG 1389  
 Db 2439 AATTAAGTGTGCAAGAAATATATGCGAAGATGTTGCTATGTAAGATTTTATAGCGAT 2498  
 Qy 1390 CTTGAAAAAGTTTCAACAGAAATTAAGTAAGTATGTAAGAAATGTAAGAAAGAAACAAA 1449  
 Db 2499 CTTAAAGATGTTAAATAATATGTTGTAAGAAATGTAAGAAATGTAAGAAAGTAAGAA 2558  
 Qy 1450 GGAAGTTTCTCAACGATGAATTTATTTCTGTGTGTAACGCAAGAAAGCCCGG 1509  
 Db 2559 TACCTTTCAAA-----TGCTTTTATCTTAAAGAACTATGTTAT 2597  
 Qy 1510 TTGCTTACACATGATCTTCAATGAATACTATCTTTTACGACAACTGATCAAAAG 1569  
 Db 2598 GGGCTTCAATGATATTTTCTTCAATCCAAAGATTAAGTTCCTTTTATGATGATCAG 2657  
 Qy 1570 CGAAGATTTCCGACAGATTAATAATGCAAGAAATGGGAGAAAGTCCAAAGATTTAGGA 1629  
 Db 2658 AGAGATTTTCCATTTGAAAGGATTTGCTTGAATTTGGGAGAGAGTGAATCAACTAGT 2717  
 Qy 1630 GAGATTCACAAAGAAATTAATGATGCGCATGTCATACCTGAGAGCAATGCAATCGCTG 1689  
 Db 2718 AGTGAATTCATTAATTAATTAAGAAAGTGAATTAATTAATTAATTAATTAATTTT 2777  
 Qy 1690 GGGACTACAGAAATTTTAAAGCAGTTTAT 1721  
 Db 2778 GACGTTACAGAAAGATTTAGAAAGTATTTT 2809

# RESULT 10 AAQ41226

ID AAQ41226 standard; DNA; 3521 BP.

XX AAQ41226;  
 AC  
 XX 27-AUG-2003 (revised)  
 DT 17-DEC-2001 (revised)  
 DT 02-SEP-1993 (first entry)  
 XX  
 DE Clone GP3 encoding major surface gp of rat P. carinii.  
 XX Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;  
 KM HIV; human immunodeficiency virus; diagnostic; PCR; ss.

XX OS Pneumocystis carinii.  
 XX  
 PN USN7958683-N.  
 XX  
 PD 01-APR-1993.  
 XX  
 PF 09-OCT-1992; 92US-00958683.  
 XX  
 PR 09-OCT-1992; 92US-00958683.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Kovace JA, Angus CW, Powell F, Edman JC;  
 DR WPI; 1992-159487/19.  
 XX  
 PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-  
 PT copy gene family - used in a vaccine and diagnostic assay for P. carinii  
 PT infection in AIDS patients.  
 XX  
 PS Disclosure; Page 37; 50pp; English.  
 XX  
 CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA  
 CC library using rat serum generated against whole rat P. carinii. Clones  
 CC reactive with polyclonal serum were evaluated to identify those encoding  
 CC gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained  
 CC ORFs encoding closely related but distinct proteins. Although none of the  
 CC clones contained the complete coding sequence, overlapping regions  
 CC allowed alignment of the three clones and generation of a putative  
 CC composite sequence encoding a protein of ca. 122 kD. PCs were used to  
 CC screen a second cDNA library constructed in modified lambda ZAP, YCDRI1.  
 CC Four positive clones, GP3, GP22, GP46 and GP14 were sequenced and were  
 CC found to contain ORFs encoding proteins similar to the original gp116  
 CC clones. The gp116 of P. carinii can be used in a vaccine against  
 CC infection in HIV-infected individuals and also as a diagnostic agent.  
 CC Study of the expression of the protein should lead to understanding its  
 CC role in the pathogenesis of P. carinii pneumonia and may lead to new  
 CC strategies for control or prevention of infection. The DNA sequence may  
 CC be used to make PCR primers for diagnostic use. NB. Due to a poorly  
 CC reproduced sequence several unreadable bases are represented by N in the  
 CC sequence below. See also AAQ41223-39. (Note: Revised entry submitted to  
 CC to correct the patent number format of US Government-owned NITs applications  
 CC to prevent clashes with ongoing US granted patent numbers. For further  
 CC information please visit the Derwent web site at  
 CC www.derwent.com/dmpi/updates/nits\_us.html.) (updated on 27-AUG-2003 to  
 CC correct OS field.)  
 XX  
 SQ Sequence 3521 BP; 1231 A; 413 C; 717 G; 740 T; 0 U; 420 Other;  
 Query Match 7.8%; Score 238.8; DB 2; Length 3521;  
 Best Local Similarity 48.6%; Pred. No. 2.3e-39;  
 Matches 1061; Conservative 0; Mismatches 1052; Indels 71; Gaps 12;  
 Qy 285 AATTAATAAATAATTCAGATTAATTAACGATTAAGATTTGCAAAAGAGATGAACAATG 344  
 Db 421 ACTTCAAAAAGTATGTAATTAATTAATAAGATGAATAATGCAAAAATATGAAGAAAATG 480  
 Qy 345 CTTATTTTGGAGGAGTATGTTCAAAAGAACTTAAGATGATTCATTAATCTTTGAGAAA 404  
 Db 481 TATACTTTTGAAGAGACGATTAATGATGTTATTAAGATTAATGATTCGATGAGGAG 540  
 Qy 405 TAACTGCTATCAAAAGAAAGCTGATTAAGTTGGCGGAAGAGTTCTTTTAAAGACATCG 464  
 Db 541 AGGATGTTACAAATTTGAAGCGTAAAGGTGGAGAGAGACTTCTTGTGAGGCGCTCGG 600  
 Qy 465 TACGATCTTAATGATCACTCATATGTGAATAAATAAATTAAGAGATTTGCCCTGTCAT 524  
 Db 601 AGGGAGTCTTAAGAAAGAACTTAATGTTAAGAAAGATTAATACGTTTGCCAGTGTT 660  
 Qy 525 GGGAGGAAAGTGAAGTAAACAACTTGTCTGTAACCAAGAAAGACATGTAAGAA 584  
 Db 661 GAGCCGAGAAAGGACGAATTAATGATGCTTTTGTGCTTGAATTCGTAATAACATGTGAGA 720

OY 585 TATTTAAATGAAAAAGTAAAGATGCGGTACTCTTAAAAAGATGTTTCAGACGACT 644  
DB 721 TCTGAAAAAAATATGGGTACTGTTTGGAGCCTTTAAAAAAGAGCTT-----AA 771  
OY 645 AGGAAGTTTAAAAAAGAAACATGCTTGAATTAATCTGACAAATCTATTTTTCATTTG 704  
DB 772 AGATTAACAAATTTAGCGGAAAGGTGCAATGAAGACTTGAAGAAATGTCATTTTTCAGGAGA 831  
OY 705 AAATTCGCGAGACGACGATATATTAATGTTATGGAATGGAGGAAATGCCAAGACA 764  
DB 832 AGCGTGTATGATG-----CGAAATGCAAGAAAGTTTGAAGGCAATGCAAGGAAA 882  
OY 765 AAATTTGCTTATATGCCACGACGACCCGATTTTGATCCAACTAGCGCAGAGCTTACAT 824  
DB 883 AAATTTATATTAAGAGCGCCAGAACTGATCTTGTCTGCAAGCCGAGGCGCTCTT 942  
OY 825 AGCAGAGATATNAGGCTGGAGAAATTTTATPAAGAGTAAAGAGATGAGATTTTAT 884  
DB 943 GTTGAGAAAGTATGGGTTGGATGATGTATTAACCGGAAAAACATGGGATTTATAT 1002  
OY 885 TGGAAA-----GAATCATCTTAAGAGATGCGACAGCTTTGTTGGCATTTGATCCAGA 938  
DB 1003 TGGAAAAATCAGAGATGATCTTCAAGAGATGCAATTTCTGCAAGATCTCTT 1062  
OY 939 TTCTAGCTTTAAAAAAGACGACAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTG 998  
DB 1063 GCTACTGTTGAGCAGAGATGAGATTAAGAGAAACAGATAAAAAGTGCACACTAAAGCGTT 1122  
OY 999 CAAAAATCCTCATGAAACATGAGGCTTTTGAAGAAATTTATGTAAGAAAAATGTTTAA 1058  
DB 1123 AGAAAAATGTATGTCTCTAAGTATTTGAATATCTAATTTGAAAAAGTTATGTAAAGATG 1182  
OY 1059 TGATGGAACGAAAAATGTGAAGATTCGCAAAATGATATTAACAACCTTGCAAAATTT 1118  
DB 1183 AAACAAAAACGAAAAATGCAAAAAATATAGATGTA--AAGAAAGATGTAACAAATCTC 1240  
OY 1119 CACTTCAAAAGTCACTAAATATGCTCTTTTGATCCAACAAAAGGAAATATGAAT--T 1176  
DB 1241 AAATTAATACTTTATCTGAAGGATGTGTCACGGAATGATGATCAAGAAATCAAGATCT 1300  
OY 1177 GTTGAATGGAGAGGTTGCCAATTTCTTACCAAGAAATGTCGAAATTTGAGTCC 1236  
DB 1301 TTATCTGAGGACACGTGCCAATTTTATTAAGAGAGTGTGCAAGATCTTGAAGTCCG 1360  
OY 1237 TATGTTTCTATTTTGAAGAAAAATGTCAG-----ATGAGAAATATGATGTAATAT 1290  
DB 1361 GAATGTTTCTATTTTGAAGAAAGCGGTGAAGATATATATGATTAAGACGTCGCAAAAT 1420  
OY 1291 ATTAAGACAACATGTTTACAAAAGAGACTTGATGCAAGGCAATTAAGTGTGCAAGAA 1350  
DB 1421 GCAGAGAGACGCTGCTATTAAGAAAGGACAAAGAGATGTTGAATTAAGTCTTTCAAAAG 1480  
OY 1351 AATATGCGAGAAATGTT--ACATGTTCAACAAAGCTGCTTGAAGAAATTTCAACA 1407  
DB 1481 GAATTAAGAGGAAACCTTGCTATGTAAAGATTTTATAGGATCTTAAGATTTGTAAGAAA 1540  
OY 1408 GAATTAAGTAAAGTATGAGAACTGAAAAAAGAAAAAAGAAAGATTTCTCAACGAT 1467  
DB 1541 TATGTGATGAAAACTGTACAAACTTGATTA----- 1573  
OY 1468 GAATTAATTAATTTCTGTGTATCAGCCAGCAAAAGAGCCCGTTTACACATGATCTT 1527  
DB 1574 AAATATCTTCCACGATGCTTTATCTTAAGAAACATGTTATAGGCTTTCAAAATGATAT 1633  
OY 1528 CGAATGAAGAACTATCTTTTACAGACAACATGATCAAAAGCGAGATTTCCCGACAGAT 1587  
DB 1634 TTTCTTCAATCCAAAGATTAAGTGGCTTTTGGATGATCAAAAGGATTTTCCATTAAA 1693  
OY 1588 AAAAATTCAGAGAAATGGGAGAGAAAGTCCAAAGATTTTGAAGAGATTTCAAAAGAAAT 1647  
DB 1634 AAGGATTTGTTGATGTAAGAGAGATGTATGAATTAAGTATGATTTCAATTAATGAAAT 1753

OY 1648 ACATGCCATGTCTATACATGAGACAGACATGCATGCTTTGGGACTACAGAAATTTTA 1707  
DB 1754 TTTGAAAAAGTATATACATGAAAAAGACGTTGTGAATCTTTTAAAGTTTCAGAGGATTT 1813  
OY 1708 AAGCAGGTTTATATGATGAACACAAAGATACCTTTGAAAAGCAAGAAAGTTGTATAAA 1767  
DB 1814 AGAAATGATTTTTTGAAGAAAAAGAGATGATTCGTTAATGACTCAGATTAATCTGACAAAG 1873  
OY 1768 TACCTAAAGAAAAGTGTAAATTAATGGTCTAAGAGAGAGATGACCGTTTCTCTTTGTA 1827  
DB 1874 GCATTTGATGAAATATGCATCAATTAATAGAGAGAGAAAGAAATTCATTAAGTGTTC 1933  
OY 1828 TGTGCTTCCAAAACGCTAGCTGTAGCTGATGTTTAAAGCGTGAAGACAGGTGCA 1887  
DB 1934 TGTGCTTTACCGAAGAAACATGTATATGTTATGTTTCCATACAGTCAAGATTTGAGT 1993  
OY 1888 GTATTTCAAAAAAATATTAAGGCTCATATATATTAATGAAATTTCTGAAAAATATACAAAT 1947  
DB 1994 AGTTTAAAGTCACATCAAGATGAAAAATTTGAAAAAATTTGAGAGAAATTTAA 2053  
OY 1948 AAATTAACAACACTGAA--AGAAATTTGCCCTTGGCATACGATTAATGCAATGATTT 2005  
DB 2054 AAACAAATATAAATGAAGCCTTGTTGAAGAACTTGACACAAATGAGGCGGACATTTGT 2113  
OY 2006 CACTTAATTTGTCAGGCTTTACGAAAGAGATAGTTGTAACAAATCAAGAGATGCTG 2065  
DB 2114 CACCACTTTATGAGAAATGTCGGA--TGACTTAAAAAAGAAAGATGCAATGCGAA 2172  
OY 2066 AGCGTTCTATTAAGAAAGGCTTGGAAAGTGTCTCAAGTGAAGTGAAGCTTCAAGAAAT 2125  
DB 2173 TGATCATTAATCTGCAAGACATCTCAAGAAATGCAATTAACCTTTGAAAAAGTTGAAAT 2232  
OY 2126 TGACTGATTA--ATCTAAATGTGAACCTGATTTGAAAGATTTGTACAGTACGGGAA 2182  
DB 2233 AGAGAGAGAGTGAATCATCTGTTGAAGGACAGTTTAAAGATATTAATGTAAGAAAGC 2292  
OY 2183 ACCTAATTAATGCTCATGCTCATGCTTTATGCAAAAGCTAACACCAAGATTAATCTGGA 2242  
DB 2293 ATTAAGAAAGGCTTGACCTGATGTTGAAG--AATGAAGCAATTCAAAATCTGTATGCTA 2351  
OY 2243 AGATGATGAGATGCTAGAAAGGAACTGTGTGAAGAAATTAAGTGAAGAAAGTGAAGAAC 2302  
DB 2352 AATGTGATGATTAATCAAGAAATATGTTGCCAAAAATTTAGTGTATTAAGTAAAAAGA 2411  
OY 2303 AGTCAAAAGCTTACCAACAGAAATTAAGACAAACCGGACGCTGATCTAAAAAAGATTA 2362  
DB 2412 GATCCCTACTTTTAAAGACGAATGGAATGCGAAAAAAGAGTTGACAAAGATGAAGA 2471  
OY 2363 AGACATATGAGAACTTAAGAAACGTGCAAGAGAAACATTAACAAGTCCAGTCTGTTT 2422  
DB 2472 ATGAGTACGATGATCTCAAAAGGCGGCAAAAAATCTACGAGGCAAGCTTAAGTTATTGC 2531  
OY 2423 TGTCACTCATTAAGAAAAAGGAA 2446  
DB 2532 TATCAAGACTTAACAAACTGTAA 2555

RESULT 11  
AAQ41230  
ID AAQ41230 standard; DNA; 3521 BP.  
XX  
XX AAQ41230;  
AC  
XX 27-AUG-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 02-SEP-1993 (first entry)  
XX  
DE Gene encoding major surface gp of rat P. carini.  
XX  
XX Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;  
KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.  
XX  
OS Pneumocystis carinii.

FH	Key	Location/Qualifiers
FT	mlec_feature	1..722 /*tag= b
FT	CDS	/note= "fragment determined by PCR" 146..3412 /*tag= a
FT	mlec_feature	626..3521 /*tag= c
FT		/note= "fragment from original GP3 clone"
XX	PN	USN7958683-N.
XX	PD	01-APR-1993.
XX	PP	09-OCT-1992; 92US-00958683.
XX	PR	09-OCT-1992; 92US-00958683.
XX	PS	(USSH ) US DEPT HEALTH & HUMAN SERVICE.
PA		Kovacs JA, Angus CW, Powell F, Edman JC;
P1		WPI: 1993-159487/19.
DR		P-PDB: AAR36706.
XX		Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
XX		copy gene family - used in a vaccine and diagnostic assay for P. carinii
XX		infection in AIDS patients.
XX		Discloure; Page 41; 50pp; English.
XX		Multiple clones were identified by immunoscreening a rat P. carinii cDNA
CC		library using rat serum generated against whole rat P. carinii. Clones
CC		reactive with polyclonal serum were evaluated to identify those encoding
CC		gp16. Three such clones (PC3, PC5 and PC14) were sequenced and contained
CC		ORFs encoding closely related but distinct proteins. Although none of the
CC		clones contained the complete coding sequence, overlapping regions
CC		allowed alignment of the three clones and generation of a putative
CC		composite sequence encoding a protein of ca. 122 kd. PCR was utilised to
CC		determine the full sequence of the gene. The 5' end of the message was
CC		identified by anchored PCR using primer JK58, which spanned the putative
CC		start codon of the composite protein. The intervening region was
CC		determined by reverse transcription followed by PCR using primers
CC		spanning the 5' end to base 722 in GP3. A single clone was identified
CC		that had an identical sequence to the first 76 bases of GP3. The
CC		complete, composite cDNA contained an ORF encoding a protein of ca. 122
CC		kd. The gp16 of P. carinii can be used in a vaccine against infection in
CC		HIV-infected individuals and also as a diagnostic agent. NB. Due to a
CC		poorly reproduced sequence several unreadable bases are represented by N
CC		in the sequence below. See also AAQ4123-39. (Note: Revised entry
CC		submitted to correct the patent number format of US Government-owned NTIS
CC		applications to prevent clashes with ongoing US granted patent numbers.
CC		For further information please visit the Dervent web site at
CC		www.dervent.com/dwpi/updates/nctis_us.html.) (Updated on 27-AUG-2003 to
CC		correct OS field.)
SQ		Sequence 3521 BP; 1238 A; 385 C; 695 G; 725 T; 0 U; 478 Other;
Query Match	6.9%; Score 211; DB 2; Length 3521;	
Beeb Local Similarity	41.7%; Pred. No. 1.3e-33;	
Matches 1030; Conservative	0; Mismatches 1357; Indels 82; Gaps 14	
Dd	1 ATGGCGGCGGCAGTCAAGCGCAGCAG--GACGACACAGAATTGATGTGAGAA 57	
Oy		
Dd	146 ATGGCACGGCCGGTTAAGAGCGACAGTCAGCAAGAGCCACAGATGAATGATGAAA 205	
Oy		
Dd	58 CATGTTTAGCTTTAATCTTAAGAAGATGAGACTAAGTGACACAGGAATGCCAAAAAA 117	
Oy		
Dd	206 CANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGAAACAABAAATGCAAAGAAGAA 265	
Oy		
Dd	118 CTAAATAATATTGCCAGAAATTGACTGAAGCAAACTAAATRTAGAACAGTAC-CAG 176	
Oy		

Db	266	CTCGAGAAATATTTGTAAGAAGTTGCAAGGAAGCAGATATAAATGTAGAGAAATGTGGATGAT	325
Oy	177	AAAACTTAAAGCTTTTCCGAAGTGGAAAAAGCAGATACAAAATGCAAAAGACTGAAGC	236
Db	326	AA	385
Oy	237	CAATATTGAGAAAAATGTACTCAATCAAAAGAAAACTTAAAGAACCAATTTAAAAAAA	296
Db	386	AAAAGTTGAGATGAAATTTAAAGAATTTTGAAGAGAACTTCAAAAATGATTTGAATAAT	445
Oy	297	AATTCAGATTATACGATTAAGATTGCAAGAGAAATGAACAATGCTATTTTTGA	356
Db	446	AA-----A-AAA	492
Oy	357	GCGAGTATGTTCAAAAGACTTAAAGTATGATGCAATCTTGAGAAATAAGTCTATCA	416
Db	493	AGACACGATTTATGATGTTATTAAGATTAACGTATGAGTTGAGGAAGATGTTCAA	552
Oy	417	AAAGAAAAGTATTAAGTGGCGAAGAAAGTCTTTTAAGAGCACTTGAGCGATCTTAA	476
Db	553	ATTGAGCGTGAATAA	612
Oy	477	TGATCAGTCACTATGTGAAAAAAACTTAAAGATTTGCCCTGTCTCATGGGAGGAAAG	536
Db	613	AGAGAAGCTTAATGTAAGAGAAAGATGAATACTGTTGCCAGTGTGAGCGAGAAAG	672
Oy	537	TGATGAGTTAACAACTTGCTGTCTGAACCAAGAAAGACATGTAAAGATTTTAATTGA	596
Db	673	CGACGAATTGATGTCCTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	732
Oy	597	AAAGATTAAGAGCGCGTACTCTTAAACAGATGTTCCAGCACACTAGAGAACTTTTAA	656
Db	733	ATTGGGTACTGTTTTCAGACCTTTAAAAAAGACCTTAAGATTAACGAATTAG-----	785
Oy	657	AAAGAAACAATGCTTGAATTAATCGAACATGCTATTTTTCACATGGAAATTCGAGAA	716
Db	786	--CGAAAAAGTCAATGAAGACTTGAAGATGTATTTTANNNNNNNNNNNNNNNNNNNN	843
Oy	717	CGACGATTAATTAATGTATTGAATTGGAGGAAAAATGCCAAGACAAATATTGCTTA	776
Db	844	TG-----CGAAATGCAAGAAAGTTTGAGACATGCAAGGAGAAAAATATTATATA	894
Oy	777	TATGCCACAGAGACCCATTTTGAATCCACTAGGCCAGAGGCTCAATPAGCAGAGATAT	836
Db	895	TAAAGCCGCAAACTCTATCTTACTGTCACAGCCAGAGGCGTCTTGTTNNNNNNNN	954
Oy	837	AGGCGTGAAGAGTTTATTAAGAGTAGAGAGATGAGAGTTTATTTGAAAGAAATC-	895
Db	955	NNNNNGATGATGTGATTAATAAAACGGGAAAAACAATGGATTATTTTGGAAAAATCAG	1014
Oy	896	-----ATCTAAGATGCGACAGCTTTGTTGGCAATTTGTATCCAAAGATTTACTTTAA	950
Db	1015	AGTGAATCTACCAAGGAAGTCAGGTACAATAATTCGTGAAGATCTTTNCTAANNNNNN	1074
Oy	951	AAAAAAGACGACAAAGAAATCCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCTCA	1010
Db	1075	NNNNNNTGAGATTAAGAGAACCAAGATTAATAAGTGCATTAAGCGTTAGAAAAATGTGA	1134
Oy	1011	TGAACATGAGCCTTAGAAAGTTATGTATTAAGAAAAATGTTTAATGATATGAGACGAA	1070
Db	1135	TGCCTTAAGATTTGAATCTGAATTTGAAAACTTATGTAAAGATGAAAAAABABNNNN	1194
Oy	1191	AAAAATGAGAAATGGCAAAATGATATTAACAAACTTGCAAAATTTTCACTTCAAAAGT	1130
Db	1195	NNNNNNCAAAAAATATTAAGTGA--AAGAAAGATGTACAAATCTCAATTAATAACTT	1252
Oy	1131	CACATATATATGCTCTTTTATCCAAACAAAAGAAATATATGAATTTGTGTATGGGAAG	1190
Db	1253	TATCTGAAGATGTCTACGGAATATGATGATCAAAATCAGATCCCTTAATCGTGNNNN	1312
Oy	1191	GTTGCGAACATTTCTTATGCAACGAAGTTGGCGA-AATTGAATCTCATTTGTTCTATT	1249
Db	1313	NNNNNNNAACCTTTTTTATTAAGAGAGATGTGAGAACTTGAAGTCGGAATGTTTCTAAT	1372

```

QY 1250 TTGAAAAAAATGTCAGATGAGAAAAATGCAATGTAATAATAAGACCAATCATGTTACA 1309
    |||||
Db 1373 TAGAAAAAGCGCTGTAAAGAT-----AATTAATTTGATTAAGCGTGCCTCAAAATGACAGA 1425
QY 1310 AAAAGAGACTTGATCAGCGGCAATAAAGTCTGCAGAAAAATATGAGAGAAATGTTAC 1369
    |||||
Db 1426 GCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1481
QY 1370 ATGTTTCAACAAAGCTGCTGTTGAAAAGTTTCAACAAGAAATAGTAATAAGATGTGAGA 1429
    |||||
Db 1482 --AATTGAAGGGAAGACTTGTCATGTAAAGTTTATATGCGATCCTTAAGTTGTAABAAA 1539
QY 1430 AACTGAAAAAAGAAAACAAAGAACTTCTCAACGATGAATTTATTTATCTGTGTAC 1489
    |||||
Db 1540 TATGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1595
QY 1490 AGCCAGCAAAAGCAGCCGGTTGCTTACACATGATCTTCGATGAAACTATCTTTTAC 1549
    |||||
Db 1596 ATCTTAAGAACTATGTTATGCGCTTGAATGATATTTTCTTCAATCCAAAGAGTTAA 1655
QY 1550 GACAACAACGTGGATCAAAAGCAGATTTCCGACAGATTAATAATGCAAGAAATGGGGA 1609
    |||||
Db 1656 GTGGCTTTTGGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1715
QY 1610 GAAAGTCCAAAGATTAGAGAGATTCCAAAAGAAATTAATCATGCGCATGTCAATACATGG 1669
    |||||
Db 1716 AGAAGTGATGAAGCTTAGTAGTATTTGAAATTTAGAAAGNGTAAACANNNN 1775
QY 1670 AGCAGCATGCAATCGCTTGGGGACTAGCAAAATTTTAAAGAGGTTTATTTGATGTAAC 1729
    |||||
Db 1776 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1835
QY 1730 ACAAAGATACTTTGAAGACCAAGAAAGTTGTGTAATAATACCTTAAGAAAGAAAGTAAATA 1789
    |||||
Db 1836 AGGATGATTCGTATATGACTAGATTAAGTAAAGGCAATGCAATGAAAGAAATCCATC 1895
QY 1790 AATGCTTAGAAGAGAGATGACCGTTCTCTTTTGTATGTGTCTTCCAAAACGCTACGT 1849
    |||||
Db 1896 AATTATATAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1955
QY 1850 GTGAGCTATGCTTAAGACGTAAGACAGGTGGAAGTATTCGAATAAAAAAATAAAG 1909
    |||||
Db 1956 GTAGTTATATGATATTCATACAGATCAAGATTGATGATTAAGTCAACATCAAGAA 2015
QY 1910 CTTCATATATTTATGTAATTTCTTGAATAATATACAAATTAATAACAACATGGAAG-- 1967
    |||||
Db 2016 ATGAAAAAATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2075
QY 1968 -AATTTGCTCTTGGCATACGATATTCATATGATTTTACCTTAATTTGTCAGGCTTTA 2026
    |||||
Db 2076 TTGTTTGAAGAACTCTGACAAACATGGGCGCGCATTTGTCACCACTTATGAGAAATG 2135
QY 2027 C-----GAAGAGAAATAGTTGACAAAAATCAAGAGATGATGAGCGCTTCTAATA 2077
    |||||
Db 2136 CCGGATGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2195
QY 2078 AAAAGAAAGCCTTGGAGATGCTCTCAAGTAGAGCTTCAAGSAAATTTGACTGATAAT 2137
    |||||
Db 2196 CAAGAAAAATGCAATTAACCTTTGAAAGTTGAATTAAGAGAGAGAGAGAGAGATGATCG 2255
QY 2138 CTAAATGTGAACCTGCAATTTGAAGAGATTTGACATGACGGGAAACGTAATAATGCGT 2197
    |||||
Db 2256 TTGAAGGCAATTTTAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2315
QY 2198 CAATAGTGGCTTATGCAAAAGCTAACCAAGATTAACCTGGAAGAGAGATGAGGATG 2257
    |||||
Db 2316 TGGAAA-----AGAAATGAAGCATTTCAAACTCTGATGTAAATGTGATGTAATA 2366
QY 2258 CTAGAAAGAACTCTGTGGAATTTAGTGAAGAGAGAGAGAGAGAGAGAGAGATTAAC 2317
    |||||
Db 2367 CCAAGAGAAATGTGTCAAAAAATTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2426

```

```

QY 2318 CAACAGATTATAGCAACCCGACACTGATCTTAAAAAAGATTATAGACATATGAGAAC 2377
    |||||
Db 2427 AAGAGCACTGAGAGAAATCGCAAAAAAGATTGAAGAAAGATGAAAGATGATGATC 2486
QY 2378 TTAGAAGACGTGACAGAGAAAGCAATGAACAAGTCAGCTTGTGTTTGTCACTCATTAAGA 2437
    |||||
Db 2487 TCAAAAAGCGCGCGAGAAAAATCTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2546
QY 2438 AAAACGAAA 2446
    |||||
Db 2547 AACCTGTA 2555
    |||||

RESULT 12
AAQ40202
ID AAQ40202 standard; DNA; 1448 BP.
XX
AC AAQ40202;
XX
DT 25-MAR-2003 (revised)
DT 09-AUG-1993 (first entry)
XX
DE Sequence of JS7-2A2U encoding the gp116 major surface antigen of rodent
DE Pneumocystis carinii (PC).
XX
KW Pneumocystis infection; diagnosis; AIDS; vaccine; therapy; ss.
XX
OS Rattus rattus.
XX
PN W09307274-A1.
XX
PD 15-APR-1993.
XX
PF 30-SEP-1992; 92WO-US008328.
XX
PR 30-SEP-1991; 91US-00768166.
XX
PR 18-OCT-1991; 91US-00781034.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Fishman JA;
XX
DR WPI; 1993-134463/16.
XX
PT Isolated DNA encoding antigens of mammalian PNEUMOCYSTIS CARINII
PT for prevention, treatment and diagnosis of PNEUMOCYSTIS CARINII
PT infection.
XX
PS Claim 4; Fig 3; 74p; English.
XX
CC The PC DNA, antigens and antibodies can be used for early diagnosis of PC
CC infection, partic. in AIDS patients. The antigens and antibodies can also
CC be used in vaccines against PC. The antigens, antibodies and antisense
CC RNA can also be used in the treatment of PC infection. (Updated on 25-MAR
CC -2003 to correct PN field.)
XX
SQ Sequence 1448 BP; 537 A; 174 C; 377 G; 360 T; 0 U; 0 Other;
XX

Query Match 6.2%; Score 192; DB 2; Length 1448;
Best Local Similarity 54.1%; Pred. No. 8.6e-30;
Matches 565; Conservative 0; Mismatches 440; Indels 39; Gaps 7;
QY 35 CACAGAAATGACATTTGATGAGAACATGTTTATGCTTAAATTTCTAAGAGATGACTAA 94
    |||||
Db 432 CACAGATGATGATTAAGAGAACACCTTTGGCTTTCATTTGTAAGACAAACATGATG 491
    |||||
QY 95 GTGACGAGAAATGCAAAAAAATTAATAAATTTGCAAGATTTGCTGAAGCAAAAC 154
    |||||
Db 492 ATGAGAAATGAGCAAAAAAGGCTCGACGAATTTGTAAGAGATTGAAGAAAGCAAGATG 551
    |||||
QY 155 TAAATATGACAAAGTACAGAAACCTTAAGGTTTTCGAGAGATGAAAGAGAGATA 214
    |||||
Db 552 AGAATTT---CAGTGTGAATAGAAAGATTAAAGACTTTGTATGATATAAAAAACAGAGAC 608
    |||||

```

```

QY 215 CAAAATGCAAGAACTGAAAGCCATATTGAGAAAAATGTAACAATCAAGAAAGAAC 274
DB 609 AAAAAATGCAAGAACTGAAAGAAATTAATGAGAACTTTGTAACGATC 666
QY 275 TTAAGAAGCAATTAATAAAAAATTCAGATTATACGGATTAGGATGCAAGAGAAATG 334
DB 669 TTGAAGCATCGGT-----AGATGACATAGAAAGTGAAGAGATTGTATAAAAACATG 719
QY 335 AACAAACAATGCCATTTTGTGAGGAGATATCTCAAAAAAGAACTTAAGATGATTGCATA 394
DB 720 AAGAAAAATATATATCTTTTGAAGAGACGAGCCAAATAGCTTTAAGAGAACTGTCTCA 779
QY 395 CTTTGAAGAAATTAAGTGCATCAAAAAGAAACGTATTAAGTTGCGGAGAAAGTTCTTTAA 454
DB 780 AGTTGAGGGAAGATGTTAGCAATTTGAAGCTTAAGAAAGTGGCAGAGAGAGCTCTTTTGA 839
QY 455 GAGCACTTCGTAGCGATCTTAATGATCAGTCATATGTGAAAAAAAATTAAAGAGATT 514
DB 840 GGGGCTCGGAAAGGAAGCTTAAGAGAGTTAAATGTAAGCAGAGATGAAAAAGGTTT 899
QY 515 GCCTGTCAATGGGAGGAAAGTATGATGATTAACAACTTGTGTCTGTAACAGAAAGAA 574
DB 900 GCCCAGGTATTAAGCCGAGAAAGCAGCAATTTGATTTTGTGCTTGTGATTCGAGTGA 959
QY 575 CATGTAAAGATATTTTAAATGAAAGATTAAGAGTGGGTACTTTAAACAGATGTTT 634
DB 960 CGTGTGCG---CGCTGAAAAAAAATCAGAAAGATTGGCCAGCTTTTAAAGAAAGCTTA 1016
QY 635 CAGCAGCACTAGAGAAATTTTAAAGAAAGCAATGCTTGAATTAATGCAACATGCTATT 694
DB 1017 AAGATG-----GGAAATTAAGGAAAAATGTCATGAAAGACTTGAGAAATGTGCTATT 1067
QY 695 TTTTCAATGGAATTTGGGAGACACATATATTAATGATGATTTGAGAGGAAAT 754
DB 1068 TTTTAC-----GGAGAAAGCGTGTATTAACAAATGTGATGAGATTAAGATCAAT 1118
QY 755 GCCAAGAAACAATAATGCTATATATGACACAGAGCCGATTTTGAATCAACTAGGCGAG 814
DB 1119 GCGAGAAAAAAGAAATCACAATTAAG--CGCCAAATTTGATTTCTGATCTGCAAGCGGA 1177
QY 815 AGGCTACAATAGCAGAGATATAGGCGTGAAGAGTTTATTAAGAGTGAAGAGAGATG 874
DB 1178 AGAGCTGTTGTTGAGAAAGTATGGGTTGATGATGATTAAGAAAGAGCTGAAGAAAG 1237
QY 875 GAGTTTATTTGGAAGATATCATTAAGATGCGACAGCTTTGTGCAATTTGATCC 934
DB 1238 GAATTATTTATTTGAAAAATCAGAGAGTGAATCAAGAGAAAGTCAAGAAATTTCTGC 1297
QY 935 AAGATT-----CTAGTCTTAATAAAAAAAGACAGAAAGAAATGCGAAGGCCTTCA 989
DB 1298 AAGATCTCTTGTCTAGTCTTGCGAGAGATGAAATGATTAAGAGATGAGGAGAAATGCG 1357
QY 990 AAAAAAGTGAAGAAATCTCATGAACATGAGGCTTTAGAAAGTTATATGAAGAAAAATG 1049
DB 1358 AAAAAAGGTTAAAAAATGGAATCTTCTAGTATTGATATCTGATTTGATGAGGATTATG 1417
QY 1050 TTTAAGTAAATGATGAAAGAAAA 1073
DB 1418 CAAAGATGCTGATTAACAAAAAAA 1441

```

## RESULT 13

AAQ41224  
ID AAQ41224 standard; DNA; 1454 BP.

XX AAQ41224,  
AC  
XX 27-AUG-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 02-SEP-1993 (first entry)  
XX  
DB Clone PCS encoding major surface gp of rat P. carinii.

```

XX Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;
KM HIV; human immunodeficiency virus; diagnostic; PCR; ss.
XX
OS Pneumocystis carinii.
XX
PN USN7958683-N.
XX
PD 01-APR-1993.
XX
PF 09-OCT-1992; 92US-00958683.
XX
PR 09-OCT-1992; 92US-00958683.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Kovace JA, Angus CW, Powell F, Edman JC;
XX
DR WPI; 1993-159487/19.
XX
PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
PT copy gene family - used in a vaccine and diagnostic assay for P. carinii
PT infection in AIDS patients.
XX
PS Disclosure; Page 35; 50pp; English.
XX
CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA
CC library using rat serum generated against whole rat P. carinii. Clones
CC reactive with polyclonal serum were evaluated to identify those encoding
CC gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained
CC ORFs encoding closely related but distinct proteins. Although none of the
CC clones contained the complete coding sequence, overlapping regions
CC allowed alignment of the three clones and generation of a putative
CC composite sequence encoding a protein of ca. 122 kD. The gp116 of P.
CC carinii can be used in a vaccine against infection in HIV-infected
CC individuals and also as a diagnostic agent. Study of the expression of
CC the protein should lead to understanding its role in the pathogenesis of
CC P. carinii pneumonia and may lead to new strategies for control or
CC prevention of infection. The DNA sequence may be used to make PCR primers
CC for diagnostic use. See also AAQ41223-39. (Note: Revised entry submitted
CC to correct the patent number format of US Government-owned NTIS
CC applications to prevent clashes with ongoing US granted patent numbers.
CC For further information please visit the Derwent web site at
CC www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 27-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 1454 BP; 545 A; 164 C; 360 G; 385 T; 0 U; 0 Other:
Query Match 6.2%; Score 191.2; DB 2; Length 1454;
Best Local Similarity 49.9%; Pred. No. 1,3e-29;
Matches 719; Conservative 0; Mismatches 663; Indels 60; Gaps 7;
QY 307 ATAAAGGATTAAGATGCAAGAAAGATGAACAACATGCTTATTTTGAAGGAGATGT 366
DB 55 ATAAAGATGAAGATGTAAGAAACATGAAGAAAAATGATTACTTTTGAAGGAAGCAGAC 114
QY 367 TCAAAAGAACTTAAGATGATGCAATCTTTGAGAAATTAAGCTCTCAAGAAAGT 426
DB 115 CCAATATGCTTTAAGGAGAAAGTGTCAAGTGAAGGAGAGATGTCGAATGTAAGCGT 174
QY 427 GATTAAGTTCGGAAGAAAGTCTTTTAAAGACACTTGATGAGCATCTTAATGATCAGTC 486
DB 175 GAAAGAGTGCAGAGAGAGCTCTTTTAAAGGCACTCGAGAGGATGCTTAAGAAAGATGCT 234
QY 487 ATATGAAAAAATACTTAAGAGATTTGCTGTCAATGAGGAGGGAAGTATGATGAGTTA 546
DB 235 AATATGAAGAAAGATGAATTAATCTGTTGCCAGATGTTAAGCGAGAAACGACGAATTTG 294
QY 547 ACAAACTTGTGTGCAAGCCGAAAGAGACATGTAAGAAATTTTAATGAAAAAGATAG 606
DB 295 ATGACTTTTGTCTTGTATCCGAGATGCAACGTGTGAGAGAGCTGAAGAAACAAATTTGGCGGA 354
QY 607 AAGTCCGTACTCTTAAGAACAGATGTTTCAGCAGCACTAGAAAGTTTAAAGAAACA 666

```

```

Db      355 GTTTCGAACCTTTAGAAAACGAAATGGAATGAGAAA-----AGCTCAAAAAAG 402
Oy      667 TGTCTGAATTAATCTGAAACAATGCTATTTTATGATGAAAATTCGGAGACGACGATATA 726
Db      403 TGTCTGAAAAGCTTGAGAAATGTCAATTTTTCAAAAGAAAGCGTGTG-----TAAAT 453
Oy      727 ATTTAATGTATTGAAATGGAGAGAAAATGCCAAGAACAAAATATTTCTTATATGCGCACCA 786
Db      454 ACAAAATGTAAAGAGATTAAGACGAATCCGAGAAAAGAAAGATTCACATATTAAGCGCG 513
Oy      787 GGACCCGATTTTGAATCCAACTAGCCAGAGGCTAATAAGCAGAGATATAGGGCTGGAA 846
Db      514 GAATCTGATTTTATAGTCCGGTCAAGCGAAAGCGGTGTGTTGAGAAATGATGGTTGAA 573
Oy      847 GAGTTTATTAAGAAAGTAGAGAGAGATGGAGTTTATTTGAAAAGATCATCTAAGAGAT 906
Db      574 GATGTGTATTAABAAACCGGAAAABACATGGGATTTATTTGAAAATTCAGAGTGAATCTA 633
Oy      907 GCGACAGCTTTGTGGCAATGTTGATCCAAAGATTTCTAGTCTTTAAAAAAAAGACGACAA 966
Db      634 CCAAGGAAGTCAAGTACAAAATTTCTGCAAGATCTCTTGCTAGTCTTGAGCAGAGATGAG 693
Oy      967 GAGAAATGCGAAGAACCCCTTCAAAAAGCTGCAAAAATCTTCATGAACATGAGGCTTTA 1026
Db      694 AATGTATGAGGGAAGAAATGCGGTAAAGCGTTAGAGAAAATGATGCTTTAAGATTTG 753
Oy      1027 GAAAGTTTATGTAAAGAAAATGGTTTAAGTATGATGAGAACAAAATATGTAAGAAATTG 1086
Db      754 GATCTAATTTGAAAAGATTAATGCAATGATGAGAAAAGAAAAGCAAAATGCAAAATTA 813
Oy      1087 CAAAATG---ATATTAACAAAATTGCAAAATTTTCACTTCAAAAGTCACTAATATATCGT 1143
Db      814 CTAGATGTAAATGTAAAGAAAGATGTACAAAATCTCAATTAATCTTATATGTGAAGGG 873
Oy      1144 CTTTTT-----GATCCAAACAAAAGGAAATATGAAATTTGTGATGGGAAAGGTTGCCA 1197
Db      874 TTGTCTACAAAATTTGAAAAGCTAAATAATCAGATCTTTTATCGTGGGACAGCTTCCA 933
Oy      1198 ACATTTCTTACGCAAGAAATGTCGAAATTTGAGATCCTATTTGTTTCTAATTTGAAAAA 1257
Db      934 ACATTTATTTACGAAGGAGAGTGTGCAAGAACTTGATCGGATGTTTCTATTTAGAAAAAT 993
Oy      1258 AAATGTCCAGATG-----AGAAAATGATGTAAAAATATTAAGACCAATGTTCAAA 1311
Db      994 GCGTGTAAAGATTAATPAAGATTGATGAAGCATGTCAAAAATGCAAGACGCGCTATATA 1053
Oy      1312 AGAGGACTTGATGCAACGGGCAATTAAGTCGCAAGAAAATATGCCGAGAA---TGTTA 1368
Db      1054 AAGGGAACAAGAGATGTGAATTAAGTTCTTTCAAAAGAAATTAAGGGGAAATCTTGCT 1113
Oy      1369 CATGCTTCAACAAAAGCTGCTTGAAAAGTTTCAACAAGATTAATTAAGTATGTGAG 1428
Db      1114 CTGTGAAGTTTATTAAGCGATCTGAAAGATCAAAAATCTGTGTGAGAAACGTGACA 1173
Oy      1429 AAACTGAAAAAGAAAACAAAGAAAGTTTCTCAAAAGATTAATTAATTTCTGTGTGA 1488
Db      1174 AAACTTAAAGAGATGATGATACCTTTCAAAA-----TGCTCT 1212
Oy      1489 CAGCCAGAAAAGCAAGCCCGGTGCTTACATGATCTTCAATGAAAATCTATCTTTTA 1548
Db      1213 TATCTTAAAGATTTATGTTATGCGCTTTCAAAATGATTTTCTTCAATCAAAAGATTA 1272
Oy      1549 CGACAAACAATGATCAAAAAGCGAATTTCCGACAGATTAATAATTTGCAAGGATGGGG 1608
Db      1273 AGTTGCTTTTGGATGATCAAGGATTTTCCATTAGAAAAGATTTGCTTGAATTTGGTG 1332
Oy      1609 AGAAAGTGCAAGATTTAGAGAGAGATTTCAAAAAGAAATTAATGCGCATGTCATACCTG 1668
Db      1333 GAGAAGTGTGATGAATCTTATGATTCATTAATGAATTTGAAAAGATGATTAACATGG 1392
Oy      1669 GAGCAGCAATGCAATCCCTTGGGACTACGAAATTTTAAACAGAGTTTATTTGATGAA 1728

```

```

Db      1393 AAAAGACGCTGTGAATCTTTAAGCTTACAGAGGATTTAGAAAAGTATTTTAAAAA 1452
Oy      1729 CA 1730
Db      1453 AA 1454

RESULT 14
AAZ94070
ID AAZ94070 standard; DNA; 249 BP.
XX
XX AC AAZ94070;
XX AC 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE P. carinii major surface glycoprotein gene HMSGp2 conserved region.
XX
XX KM Major surface glycoprotein; MSG; HMSGp2; human; pneumonia; diagnosis; ds.
XX OS Pneumocystis carinii; sp. f. hominis.
XX PN WO200009760-A1.
PD 24-FEB-2000.
PF 17-AUG-1999; 99WO-US018750.
PR 17-AUG-1998; 98US-0096805P.
XX
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q,
XX DR MPI; 2000-206025/18.
XX P-PSDB; AAY79172.
XX
XX PT Detection of the presence of Pneumocystis carinii in specimens by
XX PT identification of major surface glycoprotein (MSG) gene sequences using
XX PT two or more oligonucleotide primers derived from human P. carinii MSG
XX PT protein encoding sequence.
XX PS Claim 27; Page 103; 110pp; English.
XX
XX CC The present sequence is that of DNA encoding the conserved C-terminal
XX CC portion (see AAY79172) of the novel Pneumocystis carinii sp. f. hominis
XX CC major surface glycoprotein (MSG), HMSGp2. The invention provides novel
XX CC genes (see also AAY79063-69) that encode human-P. carinii MSGs (see also
XX CC AAY79165-71). The MSGs include a highly conserved C-terminal region of
XX CC approximately 100 amino acids. Direct detection or amplification of human
XX CC -P. carinii MSG-encoding genes, especially by PCR using primers directed
XX CC at the conserved region of the genes, provides a sensitive and specific
XX CC technique for the detection of P. carinii, and the diagnosis of P.
XX CC carinii pneumonia, especially in biological specimens (e.g. blood,
XX CC sputum) from immunocompromised patients such as those with HIV infection.
XX CC (Updated on 15-SEP-2003 to standardise OS field)
XX
XX SQ Sequence 249 BP; 73 A; 49 C; 77 G; 50 T; 0 U; 0 Other;

Query Match 6 0%; Score 185; DB 3; Length 249;
Best Local Similarity 83.9%; Pred. No. 1.5e-28;
Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Oy      2821 GAATGCAAAATCCTTAACAGACAACAGACATGGGTTACACAGATCGACACACAAGC 2880
Db      1 GAGTGCCAAATCTTCGACAGACAGACACGCTGGGTCAAAAGAGCTGACCACTACTAGC 60
Oy      2881 AGCTTACTATACATCTTACATCATCAATCAAAAATTAACATTGACATCAAGAGCGGATGC 2940
Db      61 ACTTCTACGACTAGTCCACAGTCACATCGAAGATTAACACTACCTCAACGAGCGGTGT 120
Oy      2941 AAACCAACCAAGTGTACGACGAGAGAGAAATGATGAGAGAGACGTAAACCGAGTGA 3000

```







OY 2240 GAAAGTGTATGAGATGCTAGAAAAGAACTCTGTGAAATTAAGTGAAGAAGTGAAG 2299  
Db 1059 -TACTTAATTGTGAAGATTCCTTGGAAAAATGTTGCGCGAAATTAGTTAAAAAAGTACAGA 1117  
OY 2300 AACAGTGAAGACATTACCAACAGAAATTAGACCAACCGGCAAGCTGATCTAAAAAAAAGATT 2359  
Db 1118 AGAAATGTCTTACTTTTAAAAAGCACTGATTAAGCCGAAGATGATGACCAAGATGA 1177  
OY 2360 ATAAACATATGAGAACTTAAGAAACGTGCAGAGAAACAAATGAACAAGTCCAGCTTTG 2419  
Db 1178 AGACTGAGTACGAAAAATGCTAAAAAGCGGCGAAGAAATCTACAAACAAGCTAGCTTAT 1237  
OY 2420 TTTGTCACTCATTTAAGAAAAACGAA 2446  
Db 1238 TGTATCAAAAGTCTGAAAAAGCCGCA 1264

Search completed: October 29, 2005, 02:45:56  
Job time : 1582 secs

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	132.6	4.3	737	2	AMW34880	S40C12 AG
C 2	128.6	4.2	1811	2	CG753732	CG753732 P048-4-GG
C 3	121.4	4.0	719	2	AMW34763	AMW34763 S37FL1 AG
C 4	119.8	3.9	1531	9	CG748014	CG748014 P041-4-BB
C 5	116.8	3.8	2001	9	CL118787	CL118787 ISB1-72MK
C 6	116.4	3.8	1784	9	CL081992	CL081992 CH216-16S
C 7	115.2	3.8	1981	9	CL082000	CL082000 CH216-16S
C 8	115	3.7	1727	9	CL114085	CL114085 ISB1-59J1
C 9	114.6	3.7	1453	9	AJ591978	AJ591978 Arabidopsis
C 10	114.4	3.7	1566	9	CG757757	CG757757 P053-1-DD
C 11	111	3.6	1286	9	CL118201	CL118201 ISB1-76J1
C 12	111	3.6	1491	9	CG753221	CG753221 P048-2-AA
C 13	110.2	3.6	1843	9	AG435185	AG435185 Mus musculus
C 14	109.4	3.6	699	2	AMW33354	AMW33354 S20F5 AGS4
C 15	108.8	3.5	1632	9	CL082569	CL082569 CH216-16S
C 16	108.6	3.5	1377	9	CL077307	CL077307 CH216-144S
C 17	108.4	3.5	1434	9	AJ592058	AJ592058 Arabidopsis
C 18	108.4	3.5	1797	9	CL066150	CL066150 CH216-10S
C 19	108.4	3.5	2061	9	CL081983	CL081983 CH216-16S
C 20	108.2	3.5	1519	9	AG386893	AG386893 Mus musculus
C 21	107.8	3.5	1777	9	CL110208	CL110208 ISB1-53AA
C 22	107.6	3.5	1407	9	AJ592026	AJ592026 Arabidopsis
C 23	107.2	3.5	1608	9	CL118721	CL118721 ISB1-72J1
C 24	106.8	3.5	1736	9	CL066373	CL066373 CH216-10S

25	105.8	3.4	1376	9	CG747831	P041-3-B0
26	105.6	3.4	1839	9	CL109409	ISB1-51H2
27	105.6	3.4	1863	9	CL081998	CH216-165
28	105.4	3.4	1641	9	CL078555	CH216-151
29	105.2	3.4	1716	8	CC222055	CH261-11A
30	105	3.4	1594	9	CL110653	ISB1-53P2
31	105	3.4	1824	9	CL0811234	CH216-161
32	104.6	3.4	1367	9	CL082650	CH216-168
33	104.6	3.4	2030	9	CL078553	CH216-151
34	104.4	3.4	1859	9	CL083840	ISB1-2H14
35	104	3.4	1842	9	CL044158	CH216-59P
36	103.8	3.4	1416	9	CL499037	SAT1_662
37	103.8	3.4	1533	9	CL081968	CH216-165
38	103.8	3.4	1536	9	CL078538	CH216-151
39	103.8	3.4	1808	9	CL078953	CH216-154
40	103.8	3.4	1833	9	CL058635	CH216-88H
41	103.4	3.4	1506	9	CL071996	CH216-122
42	103.4	3.4	1566	9	CL082761	CH216-171
43	103.2	3.4	1276	9	CG754010	P049-2-C0
44	103.2	3.4	1654	9	CL118709	ISB1-72H1
45	103.2	3.4	1757	9	CL078645	CH216-152

## ALIGNMENTS

Post-processing: Minimum Match 0%  
Maximum Match 100%

### Listing first 45 summaries

Database : EST:

```
EST: *
1: gb_est1:
2: gb_est2:
3: gb_hlc: *
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gest1:
9: gb_gest2:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

## SUMMARIES

db	bp	LOCUS	RESULT 1	AM334880/c
		DEFINITION	SMOCC12 ACS-1 Pneumocystis carinii cDNA 3', mRNA sequence.	
		ACCESSION	AM334880	
		VERSION	AM334880.1	
		KEYWORDS	EST.	
		SOURCE	Pneumocystis carinii	
		ORGANISM	Pneumocystis carinii	
		REFERENCE	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;	
		AUTHORS	Pneumocystidaceae, Pneumocystis.	
		TITLE	1 (bases 1 to 737)	
		JOURNAL	Smullan,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,	
		COMMENT	Berman,J.C., Kovacs,J. and Cushman,M.	
			Expressed sequence tags from Pneumocystis carinii	
			Unpublished (2000)	
			Contact: Staben C	
			School of Biological Sciences	
			University of Kentucky	
			101 Morgan Building, University of Kentucky, Lexington, KY	
			40506-0225, USA	
			Tel: 606 257 2161	
			Fax: 606 257 1717	
			Email: staben@pop.uky.edu	
		FEATURES	Location/Qualifiers	
		Source	1..737	
			/organism="Pneumocystis carinii"	
			/mol_type="mRNA"	
			/db_xref="taxon:4754"	
			/lab_host="E. coli"	
			/clone_idb="ACS-1"	
			/note="Vector: Lambda ZAP II; Site_1: EcoRI, Site_2: XhoI	
			P. carinii organisms (3x10 <sup>6</sup> ) from a single rat (99-1-6,	
			sacrificed on 3/17/99) at Cincinnati VA facilities.	
			Trizol extracted RNA. Oligo dT priming, standard	
			conditions described by vendor, Stratagene. Further	
			details see www.uky.edu/Project/Pneumocystis/"	
		ORIGIN		
		Query Match	4.3%; Score 132.6; DB 2; Length 737;	
		Best local Similarity	52.3%; Pred. No. 1.7e-16;	
		Matches 396; Conservative	0; Mismatches 319; Indels 42; Gaps 3;	
Oy	1657	TGTCATACACGTGCGACGACGATGCATTCGCTTGCGGAGCTACAGAAATTTTAAACAGGTT	1716	
db	737	TGATTGAATTGAATTGTAACGCTGCAGCAATATTAGAAAGCTTCAGAGAACTCAAGACATTA	678	

QY 1717 TTATGGATGAAACACAAAGATCTTTGAAAACCAAGAAAGTTGTGTAATACCTAAA 1776  
 DB 677 TTTTAAAAAGAAAAAGATATATTAAGAAATCAAGAAATTTGAAAGCCCTCGAAA 618  
 QY 1777 GAAAAGTGTAATTAATGGCTAGAAAGAGATGACCGTTCTCTTTGTATGTCCTTC 1836  
 DB 617 GAAAAATGTGATGGATATTTTAAAAAGAGCAAACTTATTTAGTATTTCAATGCTTTTA 558  
 QY 1837 CAAAAAGCTACGTGAGCTGATGTTAAAGACGTGAAAGACAGGTGTAAGTATTCAAA 1896  
 DB 557 CAGAAAGAAATATGTGCAACTATAGCAAGAGACAGAAATCAGTGTCAATCTTAAA 498  
 QY 1897 AAAAAATTAAGCTTCATATTTATGATTTCTTGAAAAATTAATCAATTAATAACA 1956  
 DB 497 GAGAACATAGAAAAACAGAAAAATTTTGA--TAAAGAAACGCAAAAAAATGAAACA 441  
 QY 1957 ACACGTGAAAGAAATTTGCCCTTGAGCATAGTATGCAATGATTTTCACTAATGT 2016  
 DB 440 TCACCTGAAAGAAATTTGTTATTTATGAGTCCATATTTGCAATCAGCTTATGGAATTTGT 381  
 QY 2017 CCAGCTCTTACGAAAGAGATAG-----TTGTACAAAAATCAAGAG 2058  
 DB 380 CCAAGCTATTTGAAAAAGAAAGATATGACATGATGAGTTTGTGAACTTAAGAG 321  
 QY 2059 CATGCTAGCCCTTCTATTAAGAAAGCCCTTGAAAGATGCTCTCAAGTACGCTTCAA 2118  
 DB 330 AAATGTAGCTTATTTCTGGAAAAAGACCTTGATGATGATTAATCTCATTAATTTGAAA 261  
 QY 2119 GGAATATGATGATTAATCTTAATGTGAACTGCAATGAAAGATTTGTACAGTAGG 2178  
 DB 260 GGAATTTTATGATGATTAATCTTAATGTGAACTGCAATGATGATGATGATGATGATG 201  
 QY 2179 GGAACGTAAATTAATGCGTCAATCAGTGGCTTATGCAAGCTACACCAAGATTAAGT 2238  
 DB 200 AAAACGTGAACGAATTAATCTTTAAAAATCACTGTAATGATGAAAAATTAAGC--- 144  
 QY 2239 GGAAGAGATGATGAGATGCTGAAAGAACTCTGTGAAATTTGTGAAAGAAAGTGA 2298  
 DB 143 -----GTTCAAGATGAAATTTGTGAAAGATTTGTGCAAAAAATTAAGC 102  
 QY 2299 GAACGTGCAAGCATTAACCAAGATTAAGCAACGCGCAGCTGATTAAAAAAGAT 2358  
 DB 101 AAAGATGTCCAACTTAATAATTTTAAAGAAAGCGGAAAAAGAAATTAAGACAAA 42  
 QY 2359 TATAAGACATATGAGAACTTAAGAAACGTGCAAG 2395  
 DB 41 GAAAAAGATATTAATAAAAAAAAAAAAAAAAAAACTCGAG 5  
 RESULT 2  
 CG753732/c 1811 bp DNA linear GSS 24-OCT-2003  
 LOCUS P048-4-G03.ya Ppa EcoRI BAC library Pristionchus pacificus genomic,  
 DEFINITION genomic survey sequence.  
 ACCESSION CG753732 GI:37978509  
 VERSION CG753732.1  
 KEYWORDS Pristionchus pacificus  
 SOURCE Pristionchus pacificus  
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 1811)  
 AUTHORS Strinivasan,J., Sanz,W., Jeesse,T., Wiggers-Perebolte,L., Jansen,K.,  
 Buntjer,J., van der Meulen,M. and Sommer,R.J.  
 TITLE An integrated physical and genetic map of the nematode Pristionchus  
 pacificus  
 JOURNAL Mol. Genet. Genomics 269 (S), 715-722 (2003)  
 MEDLINE 22835951  
 PUBMED 12884007  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de  
 Class: BAC ends.  
 location/Qualifiers  
 1..1811  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Ppa EcoRI BAC Library"  
 /note="The library was generated by a partial digest of  
 the genomic DNA with EcoRI and cloning into the BAC  
 vector."

ORIGIN  
 Query Match 4.2%; Score 128.6; DB 9; Length 1811;  
 Best Local Similarity 35.9%; Pred. No. 1.2e-15;  
 Matches 503; Conservative 0; Mismatches 899; Indels 0; Gaps 0;

QY 72 AATCTAAGAGAGATGACTAGTGAAGCAAGATGCAAAAAAATTAATATTTG 131  
 DB 1585 AA 1526  
 QY 132 CCAAGATTGACTGAGCAAACTTAATATGAAACAGTACACAGAAACTTAAGCTT 191  
 DB 1525 AA 1466  
 QY 192 TTGCAAGATGCAAGAGCAGATACAAAATGCAAGAACTGAAGCAATTTAGAGAAA 251  
 DB 1465 AA 1406  
 QY 252 ATGTACTACATCAAGAGAACTTAAGAAAGCAATTAATAAAATTCAGATTATAC 311  
 DB 1405 AA 1346  
 QY 312 GGAATGATGCAAGAGATGAACAACATGCTATTTTGGAGGAGATATGTTCAA 371  
 DB 1345 AA 1286  
 QY 372 AGAATTAAGATGATTGCAATCTTTGAGAAATTAAGTCTATCAAGAAACGTATAC 431  
 DB 1285 AA 1226  
 QY 432 AGTTCGCAAGAGATCTTTTAAGACACTTCGTAGCGATCTTAATGATCATATG 491  
 DB 1225 AA 1166  
 QY 492 TGAATAAACTTAAGAGATTTGCCCTGTCAAGGAGGAAAGTGATGATTAACAA 551  
 DB 1165 AA 1106  
 QY 552 CTGTGTGTAAGCAAGAAAGACATGATATTTTAATGAAAGATTAAGAGATG 611  
 DB 1105 AA 1046  
 QY 612 CGGTACTCTTAACAGATGTTTCAGCAGCACTAGAAAGTTTAAAGAAACATGCT 671  
 DB 1045 AA 986  
 QY 672 TGAATTAATGCAACATGCTATTTTACATTTGGAATTTGCGAGACGATTAATTA 731  
 DB 985 AA 926  
 QY 732 ATGTATTGAATTGGAGGAAATGCAAGAACAAATTTGCTTATATGACACGAGAC 791  
 DB 925 AA 866  
 QY 792 CGATTTTGATCAACTAGCCAGAGGCTACATATGACAGAGATATAGGCTGAGAGATT 851  
 DB 865 AA 806  
 QY 852 TTAATAGAGGTAGAGAGATGAGATTTTATTTGAAAGAAATCATCTTAAGAGATGAC 911

```

Db      805 NAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 746
Qy      912 AGCTTTGGCATGTTGATCCAGATTCAGCTTAATAAAAAAAAAAGACCAAGAGAA 971
Db      745 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 686
Qy      972 ATGCGAAGAGCCCTTCAGAAAAAGCTGCAAAATCTCATAGACATGAGGCTTAGAAG 1031
Db      685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 626
Qy      1032 TTTATGTAGAAAAATGTTTAACTATGATGAAAGCAAAAAATGTGAAGATTGCAAA 1091
Db      625 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 566
Qy      1092 TGATATTACAAACTTGCAAAATTTTCATCTTCAATGCTATATGCTTTTGA 1151
Db      565 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 506
Qy      1152 TCCACACAAAGGAATATATGAAATTTGTGATGGAGGCTTGCCACATTTCTTAGCAA 1211
Db      505 NAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 446
Qy      1212 CGAAGATTGTGCGAATTGAGTCTATTGTTCTATTGAAAAAAATGTCCAGATG 1271
Db      445 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 386
Qy      1272 AGAAAAATGCATGTATAATATAGAGCAATGTTTCAAAAGAGAGCTGTATGACGGC 1331
Db      385 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 326
Qy      1332 AAATTAAGTGTGTCGAGAAATATGCGAGAACTTTCATGCTTCAACAAAGCTGGCT 1391
Db      325 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 266
Qy      1392 TGAAGATTTCACAGAAATTAAGTAATGTGAGAACTGAAAAAGAAAACAAAG 1451
Db      265 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 206
Qy      1452 AAGTTTCTCAACAGCATGATTA 1473
Db      205 AAAAAAAAAAAAAAAAAAAAAA 184

```

RESULT 3  
 AM334763/c 719 bp mRNA linear EST 31-JAN-2000  
 LOCUS S37F11 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.  
 DEFINITION AM334763  
 ACCESSION AM334763.1 GI:6831029  
 VERSION  
 KEYWORDS EST.  
 SOURCE Pneumocystis carinii  
 ORGANISM Pneumocystis carinii  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.

REFERENCE 1 (bases 1 to 719)  
 Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Steben,C.,  
 Eiman,J.C., Kovacs,J. and Gubillon,M.  
 Expressed sequence tags from Pneumocystis carinii  
 Unpublished (2000)  
 CONTACT: Steben C  
 SCHOOL of Biological Sciences  
 University of Kentucky  
 101 Morgan Building, University of Kentucky, Lexington, KY  
 40506-0225, USA  
 Tel: 606 257 2161  
 Fax: 606 257 1717  
 Email: steben@pop.uky.edu.

FEATURES  
 SOURCE  
 1..719  
 /organism="Pneumocystis carinii"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4754"

/lab\_host="E. coli"  
 /clone\_id="AGS-1"  
 /note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;  
 P. carinii organisms (3x10e9) from a single rat (99-1-6,  
 sacrificed on 3/17/99) at Cincinnati VA facilities.  
 Trizol extracted RNA. Oligo dt priming, standard  
 conditions described by vendor, Stratagene. Further  
 details see www.uky.edu/Project/Pneumocystis/"

Query Match 4.0%; Score 121.4; DB 2; Length 719;  
 Best Local Similarity 55.3%; Pred. No. 3,4e-14;  
 Matches 289; Conservative 0; Mismatches 216; Indels 18; Gaps 2;

```

Qy      416 AAAAAAGATGATTAAGTTGCGAAGAGCTTTTAAGAGCACTTGATGATCTTA 475
Db      539 AAGCGAGCGTCATAATCTGCGATGATAATCTTTTAGAGCTTTAAAGAACAGTTA 480
Qy      476 ATGATCAGTCATATGTGAAAAAAACCTTAAGATTTGCCCTGATGGAGGAAA 535
Db      479 ATGACAGGATTAATGTAAAGAAAAAGATGAAGATTGTGCCAATTTAATAGAGGA 420
Qy      536 GTGATGATTAACTTGTGTCTGAACGAAAGAGACATGTAAATTTTAATTG 595
Db      419 GTGCGAGTTAATGTTTTCATGTATTATTTAACCTAGCTGAAAAATCTGAAAAA 360
Qy      596 AAAAAATTAAGATGCGGCTCTTAAACAGATGTTTCAGCAGCATAGGAAGTTTA 655
Db      359 AAAAAATGATTTTGTGAACTTTTAAGACAA-----ATTAGTGAATGAT 309
Qy      656 AAAAAAGATGCTTTGAATTACTCGAACATGCTATTTTTCATTTGAATTCGAG 715
Db      308 TAGAGGAAGATGTCATGAAGAGTTGAAAAATGTTATTTTAACTCAAGTG----- 254
Qy      716 ACGACATTAATTAATGATTTGAATTTGGAGAAAAATGCCAACAATAATTTGCT 775
Db      253 -----CCCTAATACAGATGATTAATTTAAAAAGCATGAGGAAAAAATTTTCA 198
Qy      776 ATATGCGACGAGGCCGATTTTGATCCCACTAGCGAGGCTACATAGCAGAGTA 835
Db      197 ACAAGCTCCAGAGATCAGATTTTCTCCAGTTCAACCAAAACAACTTGTTAAGAGA 138
Qy      836 TAGGCTGGAAGATTATTAAGAGGTAGAGAGATGAGATTTTATTGGAAGAATC 895
Db      137 TTGCTCGAAGAAATTTAATAAGAGAGAAATAACGGAATATTGTGGGAATCAG 78
Qy      896 ATCTAAGATGCGACAGCTTTGTTGCAATTTGATCCAGA 938
Db      77 GAGAAAAAGATTTGGAGATATATCTATTATTATGATGATAAAA 35

```

RESULT 4  
 CG748014 1531 bp DNA linear GSS 24-OCT-2003  
 LOCUS P41-4-B02.za Ppa Ecobi BAC Library Pristionchus pacificus genomic.  
 DEFINITION CG748014  
 ACCESSION CG748014.1 GI:37968940  
 VERSION  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 1531)  
 Srinivasan,J., Sim,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,  
 Buntjer,J., van der Meulen,M. and Sommer,R.J.  
 An integrated physical and genetic map of the nematode Pristionchus  
 pacificus  
 Mol. Genet. Genomics 269 (5), 715-722 (2003)  
 JOURNAL MEDLINE  
 22835951  
 12884007  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology

Max-Planck-Institute for Developmental Biology  
 Speemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.bommer@uebingen.mpg.de  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers

1..1531  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="ppa EcORI BAC library"  
 /note="The library was generated by a partial digest of  
 the genomic DNA with EcoRI and cloning into the BAC  
 vector."

## ORIGIN

Query Match 3.9%; Score 119.8; DB 9; Length 1531;  
 Best Local Similarity 38.6%; Pred.No. 7.7e-14;  
 Matches 523; Conservative 0; Mismatches 832; Indels 0; Gaps 0;

```

Oy 93 AAGTACGAGGATGCAAAATTAATATGTCAGGAACTTGAAGGAGAA 152
Db 177 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 236
Oy 153 ACTAATATAGACACAGTAAAGGTTTTCGAGATGGAAGAGAGA 212
Db 227 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 296
Oy 213 TACAAATGCAAGAACTGAAAGCCATTTTGAGAAAAATGCTACATCAAGGAA 272
Db 297 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 356
Oy 273 ACTTAAAGAGCAATTAATAAATTCAGATTTAAAGGATTAAGATTCAGAGAGA 332
Db 357 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 416
Oy 333 TGAACACAGATGCCATATTTTGGAGAGATATGTTCAAAAGACTTAAAGATGATGCA 392
Db 417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 476
Oy 393 TACTTGAGAAATAGTGTATCAAAAGAACTGTATAAGTTCGAGAGAGATCTTT 452
Db 477 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 536
Oy 453 AAGACACTTGTAGCGATCTTAATGATCACTCATGTGAAAAAATCTTAAGAGAT 512
Db 537 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 596
Oy 513 TTGCCCTGTATGGGAGGGAAGTATGATGATTAACTTGTGTGTAACAGAGAGA 572
Db 597 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 656
Oy 573 GACATGTAAATATTTTATTTGAAGATTAAGAGTCGCTACTCTTAACAGATGT 632
Db 657 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 716
Oy 633 TTCACAGACATGAGAGGTTTAAATAAGAACTGTCTTGAATTAATCAATGCTA 692
Db 717 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 776
Oy 693 TTTTATCATTTGAAATTTGCGAGACAGATTAATTAATGATTTGAATTTGGAGAGA 752
Db 777 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 836
Oy 753 ATGCCAAGACAAATTTGCTTATATGCAACAGAGCCGATTTTGATCCAATGAGCC 812
Db 837 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 896
Oy 813 AGAGGCTACATATGCGAGATATAGGCTGGAAGGTTTAAAGAGGTAGAGAGAGA 872
Db 897 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 956

```

```

Oy 873 TGAAGTTTATTTGGAAGAAATCATCTTAAGATGCGACAGCTTTGTCATTTGTCAT 932
Db 957 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1016
Oy 933 CCAAGATTTCTAGCTTTAAATAAAGACGACAAAGAGAAATGCGAGAGCCCTTCAGAA 992
Db 1017 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1076
Oy 993 AAGCTGCAAAATTCCTCATGACATGAGGCTTTTGAAGATTATGTAAGAAAAATGCTTT 1052
Db 1077 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1136
Oy 1053 AAGTAATGATGGAACGAAAAATGTGAAGATTGCCAAATGATATTAACAAACTTGCAG 1112
Db 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1196
Oy 1113 AATTTCACCTCAAAAGTCATTAATATCGTCTTTTGATCCACAAAGGAAATTAATGA 1172
Db 1197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1256
Oy 1173 AATTGTTGATGGAGAGGTTTCCACATTTCTTTCAGCAAGAGATTGTGCAAAATTTGA 1232
Db 1257 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1316
Oy 1233 GTCTTATTTGTTCTATTTTGAATAATGTCCAGATGAGAGAAATGATTAATAATAT 1292
Db 1317 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1376
Oy 1293 AAGACACATCTTACAAAGAGAGACTTGATGACGCGGCAATTAAGTCTGCAAGAGA 1352
Db 1377 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1436
Oy 1353 TATCGAGATATGTTACATGTTCAAACTAAAGCTGCTGAAAGTTTCAACAGAAAT 1412
Db 1437 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1496
Oy 1413 AGTAAAGTATGTGAGAACTGAAAAAGAAACA 1447
Db 1497 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1531

```

RESULT 5  
 Cui18787  
 LOCUS  
 DEFINITION  
 ISB1-72M23.T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72M23,  
 genomic survey sequence.  
 ACCESSION  
 Cui18787  
 VERSION  
 Cui18787.1 GI:40612422  
 KEYWORDS  
 GSS.  
 ORGANISM  
 Xenopus tropicalis (western clawed frog)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 Xenopodinae; Xenopus; Silurana.  
 REFERENCE  
 1 (bases 1 to 2001)  
 Kremiczki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,  
 Mardis,E. and Wilson,R.  
 A physical map of the xenopus tropicalis genome  
 Unpublished (2003)  
 JOURNAL  
 TITLE  
 CONTACT  
 Richard K Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@wustl.edu  
 Insert Length: 75000 Std Error: 0.00  
 Seg primer: T7 TAATACGACTCACTATAGG  
 Class: BAC ends  
 High quality sequence start: 384  
 High quality sequence stop: 783.  
 FEATURES  
 source  
 1..2001  
 /organism="Xenopus tropicalis"  
 /mol\_type="genomic DNA"



ORIGIN BAC library"

Query Match 3.8%; Score 116.4; DB 9; Length 1784;  
 Best Local Similarity 44.3%; Pred. No. 3.9e-13;  
 Matches 609; Conservative 0; Mismatches 758; Indels 9; Gaps 3;

Oy 78 AAAGAGATGAGCTAGTGAAGCAGAAATGCAAAAAAACTAAATAATTTGCCAAGA 137  
 Db 230 AA 289  
 Oy 138 ATTGCTGAGCAAACTAATATATAGAACAGTACAGAAAACTTAAGGTTTTTGGCA 197  
 Db 290 AAAAAAAAAAAAAAAAAAGAAA 349  
 Oy 198 AGATGGAAGAGAGATACAAAATGCAAGAACTGAAGCCATATTGAGAAAAATGTAC 257  
 Db 350 AGAAG 409  
 Oy 258 TACAACTCAAGAGAACTTAAAGACAAATTAAATAAAATTCAGATTATACGATPA 317  
 Db 410 AAGAAAAA 469  
 Oy 318 GGATTGCAAGAGATGAGCAACATGCTATTTTGGAGGAGTATGCAAAAGAACT 377  
 Db 470 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAG 529  
 Oy 378 TAAAGATGATTGCAATCTTGAAGAAATAGTCTATCAAAAGAAACGTATTAAGTTGC 437  
 Db 530 AAAAAAAAAAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589  
 Oy 438 GGAAGATGTTCTTTAAGAGCCTTCGACCATCTTAATGATGATGATGATGAGAA 497  
 Db 590 AAG 649  
 Oy 498 AAACTTAAAGATGTCCTGCTGATGGAGAGAGAGAGAGAGAGAGAGAGAGAG 557  
 Db 650 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 709  
 Oy 558 TCTGAACAGAGAGAGACATGTAGAAATTTTAAATTGAAAGATTAAGAGTGGCGTAC 617  
 Db 710 AATTAATTAATTAATTA 762  
 Oy 618 TCTTAAACAGATGTTTCAAGCCTGACGAGAGAGAGAGAGAGAGAGAGAGAG 677  
 Db 763 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 822  
 Oy 678 ACTGAGCAATGCTATTTTTCATTTGAATTCGAGAGAGAGAGAGAGAGAGAGAG 737  
 Db 823 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 882  
 Oy 738 TGAATTTGGAGAGAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 797  
 Db 883 AAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 942  
 Oy 798 TGATCAACTAGGCGAGGCTTACATATGACAGAGATATAGGCGCTGAAGAGTTTAA 857  
 Db 943 AAAAAAAAAAG 1002  
 Oy 858 GAAAGTAG 917  
 Db 1003 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1062  
 Oy 918 GTTGCAATTTGATCAAGATTTCTAGTCTTAAATAAAAGAGAGAGAGAGAGAG 977  
 Db 1063 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1122  
 Oy 978 AGAAGCCCTTCAAAAAAGCTCAAAATCTCATGACATGAGGCTTTGAAGAGTTATG 1037  
 Db 1123 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1182  
 Oy 1038 TAAGAGAAATGCTTAAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1097

Db 1183 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1242

Oy 1098 TAACAAAACCTGCAAAATTTTCACTTCAAAAGTCATATATATGCTTTTGAATCCAC 1157  
 Db 1243 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1302

Oy 1158 AAAAGAAATTAATTAATTTGTTGATGAGAGAGGTTGCCAATTTCTTACGACAGAA 1217  
 Db 1303 AAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1361

Oy 1218 TTGTGCAAAATTTGAGTCTTATTTTCTATTTTGAATAAAATGTCAGATGAGAGAA 1277  
 Db 1362 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1421

Oy 1278 TGCATGTAAATTAATTAAGAGACATGTCTCAAAAGAGAGAGAGAGAGAGAG 1337  
 Db 1422 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1481

Oy 1338 AGTCTGCAAGAAATTTGCGAGAGATGTTACATGTTTCAAAACAAAGCTGCTGAAA 1397  
 Db 1482 A-TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1540

Oy 1398 GTTCAACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1453  
 Db 1541 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1596

RESULT 7  
 CL082000/c 1981 bp DNA linear GSS 31-DEC-2003  
 LOCUS CH216-165P18 RM.1 CH216 Xenopus tropicalis genomic clone  
 DEFINITION CH216-165P18, genomic survey sequence.  
 ACCESSION CL082000  
 VERSION CL082000.1 GI:40537913  
 KEYWORDS GSS.  
 SOURCE Xenopus tropicalis (western clawed frog)  
 ORGANISM Xenopus tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus; Silurana.  
 REFERENCE 1 (bases 1 to 1981)  
 Kremlitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,  
 Mardis, E. and Wilson, R.,  
 A physical map of the xenopus tropicalis genome  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Richard K Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@wustl.wustl.edu  
 Insert Length: 175000 Std Error: 0.00  
 Seq primer: RM4 ctcaaggcgatcgctcgagc  
 Class: BAC ends  
 High quality sequence start: 265  
 High quality sequence stop: 497.

FEATURES  
 Location/Qualifiers

1..1981  
 /organism="Xenopus tropicalis"  
 /mol\_type="genomic DNA"  
 /strain="Nigerian frog"  
 /db\_xref="taxon:8364"  
 /clone="CH216-165P18"  
 /sex="male"  
 /cell\_line="Stock 248 F7A2, inbred N7"  
 /clone\_id="CH216"  
 /note="Vector: pTRABAC2.1; CHORI-216 Xenopus tropicalis  
 BAC library"

ORIGIN

Query Match 3.8%; Score 115.2; DB 9; Length 1981;  
 Best Local Similarity 44.0%; Pred. No. 6.9e-13;  
 Matches 596; Conservative 0; Mismatches 748; Indels 12; Gaps 2;

Oy 104 AATGCAAAAAAAAACTAATAATTTTGCCAGAAATGATGAGAGAGAGAGAGAGAT 163



Db	1700	ACTAAAAAAAAAGAAAAATATTAATAATCTTAAAAAAAAAAAAAAAAAAAAAAAAATCAAAAG	1641
Qy	164	AACAAGTACAGAAAACTTAAAGGTTTTGGAGAGATGAAAAAGCAGATCAAAATGCA	223
Db	1640	AAATTAATATATATATAAATTAATAATATCAATATATTAATAAAAAAAAAAAAAAAAAAACA	1581
Qy	224	AAGAATCGAAAGCCCAATTTTGAGAAAAATGTACTACATCAAGGAAACTTAAAGAG	283
Db	1580	AAATATATTAATAAATTAATTAATAAATAATATATTAATAAATAAATAAATAAATAA	1521
Qy	284	CAATTAATAAATAAATAATTCAGATTATTAAGGGAATTAAGATTGCAAGAGATGACAAAT	343
Db	1520	TAAAAAAAAATTAATAAATAATTAACATCAATTAATAAATAAATAAATAAATAAATAA	1461
Qy	344	GCTATTTTTGGAGGAGTAGTGTCAAAAGAACTTAAAGATGATTGCA-----ATACTT	397
Db	1460	ATATATTAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1401
Qy	398	TGAGAAATTAAGTCTATCAAAAGAAAGTGATTAAGTGGGAGAGACTTTTAAAG	457
Db	1400	TAAAAAATAAATAATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAA	1341
Qy	458	CACCTCGTAGGATCTTAAATGATCAGTCATGTGGAATAAATAAATAAAGATTGACC	517
Db	1340	TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1281
Qy	518	CTGTCAATGGGAGGAGAAAGTGATGATTACAAACTGTGTGTAACAGAAAGACAT	577
Db	1280	AAAAATTAACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1221
Qy	578	GTAAGAATATTTTAAATTGAATAAGTAAGAGTGGGACTCTTAAACAGATGTTGAG	637
Db	1220	AAATTAATTAATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1161
Qy	638	CAGCATTAAGAACTTTTAAAAAGAAACATGCTTGAATTACTGAAACAATGCTATTTT	697
Db	1160	TAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1101
Qy	698	ACATTGGAAATTCGGAGAGCAGCATTAATTAATTAATGATTGAATTGGAGGAAAAATGCC	757
Db	1100	AAAAAAAAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1041
Qy	758	AAGAACAATAATTCCTTATATATGCCACGAGACCCGATTGGATCAACTAGGCCAGAG	817
Db	1040	AAAAAAAAAAAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	983
Qy	818	CTACAAATGCAAGATATAGGCGCTGGAAGCTTTATTAAGAAAGTAAGAGAGATGCGAG	877
Db	982	----TATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	927
Qy	878	TTTTTATGGAAGAAATCATCTAAGAGATGCGACACTTTGTGTCATTGTTGATCAAG	937
Db	926	AAAAAAAAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	867
Qy	938	ATTCTAGTCTTAAATAAATAAAGCAGCAAGAGAAATGCGAGAGACCCTTCAAAAAGCT	997
Db	866	AAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	807
Qy	998	GCAAAAAATCCTCATGACATGAGCTTTAGAAAGTTTATGTAAAGAAAAATGCTTTAAGTA	1057
Db	806	ACAAAAATTAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	747
Qy	1058	ATGATGCAAGCAAAAAATGTGAAGATTTGCAAAATGATTTAACAAAACTTGCAAAATTT	1117
Db	746	TAAAAAATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	687
Qy	1118	TCACCTCAAAAGTCACTAATATCGTCTTTTGTATCACAACAAGAAATATATGAATTTG	1177
Db	686	AAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	627
Qy	1178	TTGATGGAAAGGGTCCACATTTCTTAGCAAGAAATTTGCAAAATTTGAGTCTT	1237

[illegible]

RESULT	8
LOCUS	CL114085
DEFINITION	CL114085 1727 bp DNA linear GSS 05-JAN-2004
ACCESSION	ISB1-59011_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59011,
VERSION	genomic survey sequence.
KEYWORDS	CL114085
SOURCE	CL114085.1 GI:40607720
ORGANISM	GSS. Xenopus tropicalis (western clawed frog)
REFERENCE	Xenopus tropicalis Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana. 1 (bases 1 to 1727)

REFERENCE	1 (bases 1 to 1727)
AUTHORS	Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE	A physical map of the xenopus tropicalis genome
JOURNAL	Unpublished (2003)
COMMENT	Contact: Richard K Wilson

Washington University School of Medicine  
Email: submissions@wustl.edu  
Insert length: 75000 Std Error: 0.00  
Seq primer: Sp6 ATTAGGTGACACATG  
Class: BAC ends  
High quality sequence start: 280  
High quality sequence stop: 1423.

```

FEATURES
    source
        location/Qualifiers
            1..1727
                /organism="Xenopus tropicalis"
                /mol_type="genomic DNA"
                /db_xref="taxon:8364"
                /clone="ISB1-5911"
                /clone_1bp="ISB1"
                /note="Vector: pBelBAC11; ISB-1 Xenopus tropicalis BAC
                    library Segment 1"

```

	Query Match	3.7%;	Score 115;	DB 9;	Length 1727;
	Beet Local Similarity	42.3%;	Pred. No. 7.4e-13;		
	Matches 601;	Conservative 0;	Mismatch 819;	Indels 0;	Gaps 0;
Qy	34 ACACAGATGCGATGATGAGGAACATGTTTACCTTAAATTCTAAAGGAAGATGCACTA				93
Db	165 AAAAAAAAAAAAAAAAAAAAAAACTCTCTCTTAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAA				224
Qy	94 AGTGAAGCAGATGCCAAAAAAAACTAAAAAAAAATTGCGCAAGATTGACTGAAGCAAAA				153
Db	225 AAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAA				284
Qy	154 CTAATATTAGACAGTACACAGAAACTTAAAGGTTTTGCGAAGATGAAAAAGCAGAT				213
Db	285 AAAGAAAAA				344

```

Oy 214 ACAAATGCAAGAACTGAAAGCCAAATTGTGAGAAAAATGTACTCAATCAAGAGAAA 273
Db 345 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404
Oy 274 CTTAAGAAGCAATTAATAAATAATTCAGATTATACGATTAAGATTTGCAAGAGAAAT 333
Db 405 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 464
Oy 334 GAACAAACATGCTATTTTTGGAGGAGTATGTTCAAAAGAACTTAAGATGATTCAT 393
Db 465 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 524
Oy 394 ACTTGAGAAATTAAGTCTATCAAAAGAACTGTATAAGTTGGCGAAGAACTCTTTA 453
Db 525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 584
Oy 454 AGAGCACTTCGAGCATCTTAATGATCAGTCATATGTGAAGAAAACTTAAGAGATT 513
Db 585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 644
Oy 514 TGCCCTGTCATGGGAGGAGAAAGTATGATTAACTTGTGTCGACCGAGAGAG 573
Db 645 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 704
Oy 574 ACATCTAAGATATTTTAAATTAAGAAAGTGCCTGCTTAAACAGATGTT 633
Db 705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 764
Oy 634 TCAGCAGCACTAGAGATTTTAAAGAAACATGCTTGAATTTCTGCAACATGCTAT 693
Db 765 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 824
Oy 694 TTTTACATTGGAATTTGCGAGACGACATTAATTAATGATTTGGAATTGGAGAAA 753
Db 825 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 884
Oy 754 TGCCAGAGAAATATGCTTATATGCAACAGACCCGATTTGATCCAACTAGGCCA 813
Db 885 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944
Oy 814 GAGCCTCAATAGCAGAGATTAAGGCTGGAAGATTTTAAAGAGTAGAGAGAGAT 873
Db 945 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1004
Oy 874 GGAGTTTATTTGGAAGATCATTAAGAGTGCAGACCTTTGGCATTTGATC 933
Db 1005 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1064
Oy 934 CAAGATTTCTAGCTTAATAAATAAAGCAACAAAGAAATGCGAAGACCTTCAAAA 993
Db 1065 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1124
Oy 994 AGCTCAAAAAATCCTCATGAACATGAGCTTTAGAAAGTTTATGTAAGAAAAATGTTT 1053
Db 1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1184
Oy 1054 AGTAATGATGAGCAAGAAATGTGAAGATTGCAAAATGATATTACAAAATTGCAA 1113
Db 1185 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1244
Oy 1114 ATTTTCATCTCAAAAGTCAATTAATGCTCTTTTGTCCAAACAAAGAAATTAAGA 1173
Db 1245 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1304
Oy 1174 ATGTGATGAGGAGGCTGCAACATTTCTTACAAAGAAAGTTGCGAAATTTGAG 1233
Db 1305 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1364
Oy 1234 TCTATTTGTTCTATTTGAAAAAAATGTCAGATGAGAAATGCAATGTAATAATTA 1293
Db 1365 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1424
Oy 1294 AGAGCAACATGTTCAAAAGAGGACTTGATGACGCGCAATTAAGTCTGCAAGAAAT 1353

```

```

Db 1425 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1484
Oy 1354 ATGCGAGAAATGTTACATGCTTCAACAAAGCTGCTGMAAGTTTCAACAAAGATTA 1413
Db 1485 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1544
Oy 1414 GTAAAGTATGTGAGAACTGAAAAAGAAACAAAGAA 1453
Db 1545 ACAAACACAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1584

RESULT 9
AJ591978/c 1453 bp DNA linear GSS 15-JAN-2004
LOCUS Arabidopsis thaliana T-DNA flanking sequence, right border, clone
DEFINITION 598D01, genomic survey sequence.
ACCESSION AJ591978
VERSION AJ591978.1 GI:37941602
KEYWORDS GSS; right border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
1 Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
Chauvin,S., Bechtold,N., Cruaud,C., Dehose,R., Pelletier,G.,
Lepintec,L., Caboche,M. and Lecharny,A.
T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)

JOURNAL MEDLINE
2363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 1453)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publications/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.inbio.gen.fr).

FEATURES
source
1..1453
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Wassiljewskij4"
/db_xref="taxon:3702"
/clone="598D01"
/clone_id="Arabidopsis thaliana T-DNA insertion lines"
1..1453
/misc_feature
/note="T-DNA flanking sequence
right border"

ORIGIN
Query Match 3.7%; Score 114.6; DB 9; Length 1453;
Best Local Similarity 38.9%; Pred. No. 8.9e-13;
Matches 544; Conservative 0; Mismatches 852; Indels 3; Gaps 1;

Oy 79 AAGAGAGTGACCTAAGTACGACGAAATGCAAAAAAATTAATTTGCCAAGAA 138
Db 1427 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1368
Oy 139 TTGACTGAGCAAACTTAATATATGAACTACAGAAATCTTAAGTTTTCGAA 198

```

[illegible]

Oy	1279	GCAGTAAAAATTAAGAGCAACGTGTTACAAAGAGAGACTGATGACGGGCAAAATPAA	1338
Db	230	AA	171
Oy	1339	GTGCTGCAGAAAAATATGCGAGCAATGTTACATGCTTCAACACAAAGCTGGCTTGA	1398
Db	170	AA	111
Oy	1399	TTTCACACAGATTAGTAAAGTAGTGAGAACTGAAAAAGAAACAAAGAAAGTTTC	1458
Db	110	AA	51
Oy	1459	TCAACAGTCGATTATTTA	1477
Db	50	MAAAAAAAAAAAAAAAAANNNA	32

RESULT 10	CG575757/c	1566 bp	DNA	linear	GSS 24-OCT-2003
LOCUS	CG575757/c				
DEFINITION	CG575757	P0331-D07.zc	Ppa	Scori	BAC Library
ACCESSION	CG575757				
VERSION	CG575757.1	GI:37986636			
KEYWORDS	GSS				
SOURCE	<i>Pristionchus pacificus</i>				
ORGANISM	<i>Pristionchus pacificus</i>				

## REFERENCE

## AUTHORS

**TITLE**

•  
•  
•  
•  
•

JOURNAL

MEDLINE  
PITMED

**FORWARDED COMMENT**

## PREATTIRES

**BOUR**

**ORIGIN**

Cheng, M.

Query m  
Best Lo

### Matches

?

57

**D5**

?

57

D6

Qy

db

[illegible]

```

QY 258 TACAAATCAAGAAATCTTAAAGAACTTAAATTTCAAGATTATACGATTA 317
DB 1377 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1318
QY 318 GGATTGCAAGAGATGAACAATGCGCTATTTTGGAGGAGTATGTCTCAAAAGACT 377
DB 1317 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1258
QY 378 TAAAGATGATTCGCAATCTTGGAGAAATAGTCTATCAAAAGAAACGTGATTAAGTTGC 437
DB 1257 AAAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1198
QY 438 GGAAGAGATTCTTTTAAGAGCACTTCGTAGCATCTTAATGATCATGTGTGAAA 497
DB 1197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1138
QY 498 AAAACTTAAAGATTTGGCCCTGTCATGGGAGGAGAAAGTGTGATTAACAACCTTGTG 557
DB 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1078
QY 558 TCTGAACGAGAAAGACATGTAGCAATATTTTAATGAAAAAGATAAGAGTGGCTAC 617
DB 1077 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1018
QY 618 TCTTAAACAGATGTTTCAAGCACTAGCAAGTTTAAAGAAACATGCTTGAATT 677
DB 1017 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 959
QY 678 ACTGACAAATGCTATTTTACATTTGGAATTCGGAGCGACGATATTAATTAATGTAT 737
DB 958 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 899
QY 738 TGAATTGGAGAGAAATGCGCAAGAACAAATATGCTTATATGCCACGAGCCGACTT 797
DB 898 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 839
QY 798 TGAATCACTAGGCGAGGCTTACATATGACAGAGATATAGGCGCTGGAAGCTTTATTA 857
DB 838 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 779
QY 858 GAAAGTAGAGAGATGAGAGTTTATTTGGAAGATCATCTTAAGATGCGACGCTTT 917
DB 778 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 719
QY 918 GTTGCAATGTTGATCCAAAGATTCTAGTCTTAAAAAAAAAAAGACAAAGAGAAATGCGA 977
DB 718 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 659
QY 978 AGAGCCCTTCAAAAAGCTGCAAAATCTCATGACATGAGGCTTTAGAAAGTTTATG 1037
DB 658 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 599
QY 1038 TAAAGAAAAATGTTTAAATGATGAAACGAAAAATGTGAAGATTGCAAAATGATAT 1097
DB 598 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 539
QY 1098 TAAACAACTTGCATAATTTTCACTTCAAAAGTCACTATATATCGTCTTTTGTATCAAC 1157
DB 538 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 479
QY 1158 AAAAGCAATTAATGAATTTGTTGATGGAGAGGTTGCCACATTTCTTAGCAGCAAGA 1217
DB 478 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 419
QY 1218 TTGTGGAATTTGAGTCTTATTTGTTTGAATTTTGAATTTTCCAGATGAGGAGAA 1277
DB 418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 359
QY 1278 TGCATGTAAAAATATTAAGACAACTGTTCACAAAAGAGACTTGTGACGGGCAATTA 1337
DB 358 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 299

```

```

QY 1338 AGTCTGCAAGAAATATGCGAGGAATGTTACATGTTCCAAAGCTGCTTGA 1397
DB 298 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 239
QY 1398 GTTTCACAGAAATTAAGTATGATGAGAACTGAAAGAAAGAACAA 1448
DB 238 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 188

RESULT 11
CL119201
LOCUS CL119201
DEFINITION ISB1-76J12_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-76J12,
genomic survey sequence.
ACCESSION CL119201
VERSION CL119201.1 GI:40612836
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1256)
AUTHORS Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGTGACACTATG
Class: BAC ends
High quality sequence start: 233
High quality sequence stop: 402.
Location/Qualifiers
source
1..1256
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-76J12"
/clone_1lb="ISB1"
/note="Vector: pBelBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 3.6%; Score 111; DB 9; Length 1256;
Best Local Similarity 38.2%; Pred. No. 4.8e-12;
Matches 471; Conservative 0; Mismatches 763; Indels 0; Gaps 0;

QY 78 AAAGAGATGCAATAGTACGAGATGCAAAATCTAAATAATTTGCCAAGA 137
DB 23 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82
QY 138 ATTGACTGACCAAACTTAATATAGAACAGTACACAGAAACTTAAAGTTTGGCA 197
DB 83 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 142
QY 198 AGATGAAAGCAGTACAAATGCAAGACGTAATGCAAGCCATTTTGGAAAAATGTAC 257
DB 143 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 202
QY 258 TACATCAAGAGAAATCTTAAAGAGCAATTTAAATAATTCAGTTATACGATTA 317
DB 203 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 262
QY 318 GGATTGCAAGAGATGAACAATGCTATTTTGGAGGAGTATGTTCAAAAGAACT 377
DB 263 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 322
QY 378 TAAAGATGATTCGAATCTTTGAGAAATTAAGTGTATCAAAAGAAAGTGTAAAGTTGC 437

```









QY 589 TTAATTGAAAAAGATTAAGAGTGGCTACTCTTAAACAGATGTTTCAGCAGACTAGCA 648  
DB 1182 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAAN 1123  
QY 649 AGTTTAAAAAAGAAACATGCTCTGTAATCTCGAACATCTCTTTTACATTGGAAT 708  
DB 1122 ANANANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 1063  
QY 709 TGGGAGCGAGATATAATTGATTTGATTGGAGGAAATGCGAGAACAAAT 768  
DB 1062 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 1003  
QY 769 ATTGCTTATATGCAACGAGCCGATTTTATCCAGTACGAGGCTACATAGCA 828  
DB 1002 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 943  
QY 829 GAGGATATAGGCGTGAAGAGTTTATAGAGAGTAGAGAGATGAGTTTATTTGA 888  
DB 942 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 883  
QY 889 AAGAAATCTTAAGAGATGCGACGCTTTGTCGATTTGATCCAGATTTCTAGCTT 948  
DB 882 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 823  
QY 949 AAAAAAAGAGCAAGAGAAATGCGAGAGCCCTTCAAAAAAGCTGCAAAAAATCCT 1008  
DB 822 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 763  
QY 1009 CATGAACATGAGGCTTAGAAGATTATGTAGAAAAATGCTTAACTATGATGAGACG 1068  
DB 762 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 703  
QY 1069 AAAAAATGTGAAGATTGCAAAATGATTTAACTGCAAAATTTTCACTTCAAA 1128  
DB 702 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 643  
QY 1129 GTCACTAATAATCGCTTTTTCATCAAAAGAAATTAATGAATTTGATGGGA 1188  
DB 642 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 583  
QY 1189 GGGTTCACACATTTCTTGAACGAGATTTGTGGAATTTGAGTCTATTTCTAT 1248  
DB 582 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 523  
QY 1249 TTTGAAAAAATGTCAGATGAGAAATGATTAATTAAGGCAACATGTAC 1308  
DB 522 TTTAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 463  
QY 1309 AAAAAAGGACTTGATGACGGGCAATTAAGTGTGCAAGAAATTTATGCGAGATGTTA 1368  
DB 462 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 403  
QY 1369 CATGATTCAAACAAAGCTGCTGGAAGATTTCAACAGATTAATTAAGATTTGAG 1428  
DB 402 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 343  
QY 1429 AAATCTGAAAAAGAAACAAAGAGTTTCTCAACGATTAATTTATTTCTGTGTA 1488  
DB 342 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 283  
QY 1489 CAGCAGCAAAAGCAGCCGCTGCTTACATGATCTTCGATGAATTAATCTTTTA 1548  
DB 282 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 223  
QY 1549 CGACAACTGATCAAAAGCAGATTTCCCGACAGATTAATA 1592  
DB 222 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 179

RESULT 14  
AM33354/c AM33354 699 bp mRNA linear EST 31-JAN-2000  
LOCUS

DEFINITION S20F5 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.  
ACCESSION AM33354  
VERSION AM33354.1 GI:6829711  
KEYWORDS EST.  
SOURCE Pneumocystis carinii  
ORGANISM Pneumocystis carinii  
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 699)  
Smilian A.G., Arnold J., Weise M., Wunderlich J., Staben C.,  
Edman J.C., Kovacs J. and Cushion M.  
Expressed sequence tags from Pneumocystis carinii  
Unpublished (2000)  
JOURNAL Contact: Staben C  
COMMENT School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu.  
FEATURES  
Source location/Qualifiers  
1..699  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_lib="AGS-1"  
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;  
P. carinii organisms (3x10<sup>6</sup>) from a single rat (99-1-6,  
sacrilficed on 3/17/99) at Cincinnati VA facilities.  
Trizol extracted RNA. Oligo dT priming, standard  
details see www.uky.edu/project/pneumocystis/"

ORIGIN  
Query Match 3.6%; Score 109.4; DB 2; Length 699;  
Best Local Similarity 56.1%; Pred. No. 9.7e-12;  
Matches 317; Conservative 0; Mismatches 201; Indels 47; Gaps 4;  
QY 2551 GATGATCCGTACACAAATTAAGAGCTTAAGCATTTGATGGCAGCAAGTATTGGA 2610  
DB 681 GATGAGGTGATCAGAGTAGAGTCAAGCATTTATGCAAGAGATAGCATTTGGA 622  
QY 2611 AGATATGATTTGAAGAAAGATGTAATTAATGGAATCAGATTCAGATTAAGAG 2670  
DB 621 TTGATTTGGAATGAAGAAAGATGTAATTAATGGAATTCAGATTTTAAAGAG 562  
QY 2671 GATTGCAAGCTTAGAAGAGTATGCAAAAGATTAATTAAGGCTTGTCCAAATCTGAG 2730  
DB 561 GATTGTCCAGATTAATAAAGCTTTGGAAATTAATTAACA---CTTATGTAATCTGGA 505  
QY 2731 CCTGTGAGGTGAAGCGCA-----CGAAACATGACAGAAATGTA 2772  
DB 504 CCATTGAAATTAAGCTCTCATACAGAGAAATTAACAGAAACAGAGCAAGCAAG 445  
QY 2773 ACGCACTACACAAACAAACACCGTTGCCGATCGAAGCAAG-----GAAT 2824  
DB 444 ACGGAAACAAAGAGCAAGCAAGCTGATGCAAGCTGATGAAAGACCGTTGAGAG 385  
QY 2825 GCAATCTTACAGACACAGACATGAGTTACACAGATCGACACACAGACAGT 2884  
DB 384 ACTGTTACAGAAACCAAGTCAATGAGGTGAGCAAGTCAATGCAATGCAAGTACGA 325  
QY 2885 CTACTATCACTTACATCAATCAATCAATTAATTAATGACATCAAGAGCGATGCAAC 2944  
DB 324 CAACGAGTACGTCAACGAGTACGAGTACGATGAGTTCGATCGAAGGCAAGC 265  
QY 2945 CAACCAAGTAC-----GACAGAGAGAGATGATGACGAGAGC 2986  
DB 264 CTACCAATGTAACCAAGCATTCAGCAAGAGACAGAAAGAGAGATGATGAAGAG 205  
QY 2987 TGAACCGAGTGAAGGCGCTGAGATGATGAGTGAATGTGATGAGGCGGCTGATAG 3046

Db 204 TGAACCCGAAAGAGGAATGAAAATAGAGCTTCGTATGTGTTAAATATGTTGTTGG 145  
Oy 3047 CAATGCTATTTCTTCATGATTTA 3071  
Db 144 GAGTATTTGTTATGGGATGATGTA 120

RESULT 15  
CL082569 1632 bp DNA linear GSS 31-DEC-2003  
LOCUS CH216-167P5\_Sp6.1 CH216 Xenopus tropicalis genomic clone  
DEFINITION CH216-167P5, genomic survey sequence.  
ACCESSION CL082569  
VERSION CL082569.1 GI:40538482  
SOURCE GSS.  
ORGANISM Xenopus tropicalis (western clawed frog)  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Neobatrachia; Pipidea; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 1632)  
Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,  
Mardis, E. and Wilson, R.  
A physical map of the xenopus tropicalis genome  
Unpublished (2003)  
Contact: Richard K Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: submls@wustl.edu  
Insert Length: 175000 Std Error: 0.00  
Seq primer: Sp6 ATTTAGGTGACACTATAG  
Class: BAC ends  
High quality sequence start: 1062  
High quality sequence stop: 1114.  
Location/Qualifiers  
1. 1632  
/organism="Xenopus tropicalis"  
/mol\_type="genomic DNA"  
/strain="Nigerian frog"  
/db\_xref="taxon:8364"  
/clone="CH216-167P5"  
/sex="male"  
/cell\_line="Stock 248 F7A2, inbred N7"  
/clone\_1lb="CH216"  
/note="vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis  
BAC library"

ORIGIN  
Query Match 3.5%; Score 108.8; DB 9; Length 1632;  
Beet Local Similarity 41.6%; Pred. No. 1.4e-11;  
Matches 569; Conservative 0; Mismatches 799; Indels 0; Gaps 0;

Oy 78 AAAGAGATGAGCTAAGTGAAGAGAAATGCAAAAAAATTAATATATGCGAAGA 137  
Db 263 AA 322  
Oy 138 ATTGAAGCAAACTAATATATAGAACAGTACAGAAAACTTAAGTTTGGCA 197  
Db 323 AA 382  
Oy 198 AGATGGAAGAGAGTACAAATGCAAGAACTGAAGCAATATTGAGAAAAATGAC 257  
Db 383 AA 442  
Oy 258 TACATCAAGAGAACTTAAGAGCAATTAATAAAATTCAGATTATAGCATTA 317  
Db 443 AA 502  
Oy 318 GGATTGCAAGAGAAAGCAAACTGCTATTTTGGAGGAGTATGTTCAAGAACT 377  
Db 503 AAAAAAAAAAAAAAAAAAAAAAAAAATTTTATTAATAAAAAAAAAAAAAAAAAA 562

Oy 378 TAAAGATGATTGCAATACTTTGAGAAATAAGTCTATCAAAAGAGGTGATAAGTTGC 437  
Db 563 AA 622  
Oy 438 GGAAGAGTTCTTTAAGACACTTCGTAGCATCTTAATGATCATGTATGTGAAA 497  
Db 623 AAAAAAAAAATAA 682  
Oy 498 AAAAATTAAAGAGATTGCCCCGTCAAGGAGGAAAGTGATGATTAACTTGTG 557  
Db 683 AAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 742  
Oy 558 TCTGAACCAAGAGACATGTAGAAATTTTATTTGAATAAGATGAGCGGTAC 617  
Db 743 AAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 802  
Oy 618 TCTTAACACAGATTTTCAGACAGCACTAGAGATTTAATAAAGAAACATGCTTGAT 677  
Db 803 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATTAATAAAAAAAAAAAAAA 862  
Oy 678 ACTGAAACAATGCTATTTTACATTGGAATTGCGAGACGACATTAATTAATGTAT 737  
Db 863 AA 922  
Oy 738 TGAATTGGAGAGAAATGCCAAGAAATATGCTTATATGCGACACGACCGATT 797  
Db 923 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATTAATAAAAAAAAAAAAAA 982  
Oy 798 TGATCCAAGTACAGCAGAGCTACATAGCAGAGATATAGGCTGAGATTATTA 857  
Db 983 TTAATAATAA 1042  
Oy 858 GAAGTGAAGAGAGATGAGTTTATTTGGAAGATCATTAAGATGCGACAGCTTT 917  
Db 1043 AAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA 1102  
Oy 918 GTTGCAATTGTGATCAAGATTCTAGCTTAATAAAAAAAAAAGACAAAGGAATGCA 977  
Db 1103 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATTAATAAAAAAAAAAAAAA 1162  
Oy 978 AGAAGCCCTTCAAAAAGCTGCAAAAATCCTCATGAAATGAGGCTTTAGAAAGTTATG 1037  
Db 1163 AAAAAAAAAATAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAA 1222  
Oy 1038 TAAAGAAATGCTTAAGTATGATGAGCAAAAAATGCAAGATTGCAAAATGATAT 1097  
Db 1223 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAA 1282  
Oy 1098 TAACAAACTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTTTTGTCCAC 1157  
Db 1283 NAAAAAAAAATAATAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAAAAAAAAAAAAAA 1342  
Oy 1158 AAAAGAAATATGAAATTTGTTGATGAGGAAGGTTCCAACTTTCTTACAAAGAGA 1217  
Db 1343 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAA 1402  
Oy 1218 TTGTGCGAATGAGTCTATTTCTTATTTGAAAAAAATGTCAGATGAGAAA 1277  
Db 1403 AAAAAAAAAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAAAAAAAAA 1462  
Oy 1278 TGCATGTAAATTAAGAGCAATGTTCAAAAGGACTTGATGACGGCAATTA 1337  
Db 1463 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAA 1522  
Oy 1338 AGTCCTCAAGAAATATGCGAGAAATGTTACATGTTCAAAAGAGCTGCTGAAA 1397  
Db 1523 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAATAAAAAAAAAAAAAA 1582  
Oy 1398 GTTTCACAGAAATTAGTAAAGTATGTGAGAACTGAAAAAGAAA 1445  
Db 1583 AAAAAAAAAATAAATAATAAATAAATAAATAAATAAATAAATAA 1630

Search completed: October 29, 2005, 08:57:50  
Job time : 9463 secs

---

*This Page Blank (uspto)*



OY	301	CAGATTAAAGGATTAAGATTGGAAAAGAAATGAACAACAATCCATTATTTTGGAGGCA	360
Db	301	CAGATTAAAGGATTAAGATTGGAAAAGAAATGAACAACAATCCATTATTTTGGAGGCA	360
OY	361	GTATGTTCAAAAAGAACTTAAAGATGATTTGCATATCTTTGAGAAATTAAGTGCTATCAAAAG	420
Db	361	GTATGTTCAAAAAGAACTTAAAGATGATTTGCATATCTTTGAGAAATTAAGTGCTATCAAAAG	420
OY	421	AAACGTGATTAAGTTGGCGGAAGAAAGTTCTTTTAAAGACATTCGTAGCGATCTTAAATGGA	480
Db	421	AAACGTGATTAAGTTGGCGGAAGAAAGTTCTTTTAAAGACATTCGTAGCGATCTTAAATGGA	480
OY	481	TCAGTCATATGTGAAAAAAAACCTTAAAGATTTTCCCTGTCAATGGGAGGGAAGATGAT	540
Db	481	TCAGTCATATGTGAAAAAAAACCTTAAAGATTTTCCCTGTCAATGGGAGGGAAGATGAT	540
OY	541	GAGTTTAACAAACTTGTGTCTGACCAAGAAAGACATGTAAAGATTTTAAATGGAATAA	600
Db	541	GAGTTTAACAAACTTGTGTCTGACCAAGAAAGACATGTAAAGATTTTAAATGGAATAA	600
OY	601	GATTAAGAAAGTCCGCTACTCTTAAACAGATGTTTCAAGACGACTAGGAAGTTTAAAAAA	660
Db	601	GATTAAGAAAGTCCGCTACTCTTAAACAGATGTTTCAAGACGACTAGGAAGTTTAAAAAA	660
OY	661	GAAACATGTCTTGAAATTACCTGGAACAAATCTTATTTTACATTGGAAAAATGCGGAGAC	720
Db	661	GAAACATGTCTTGAAATTACCTGGAACAAATCTTATTTTACATTGGAAAAATGCGGAGAC	720
OY	721	GATATTAATTAATGATTAATGATTTGGGAGGAATAATGCCAAGAACAAATTAATGCTTATATG	780
Db	721	GATATTAATTAATGATTAATGATTTGGGAGGAATAATGCCAAGAACAAATTAATGCTTATATG	780
OY	781	CCACCAGAACCCGATTTTTCATCCAATAAGCCAGAGGCTCAATATAGCAGAGATATAGG	840
Db	781	CCACCAGAACCCGATTTTTCATCCAATAAGCCAGAGGCTCAATATAGCAGAGATATAGG	840
OY	841	CTGGAAGAGCTTTTATTAAGAGGTAGAGAGATGAGATTTTATTTGGAAGATTCATCTTA	900
Db	841	CTGGAAGAGCTTTTATTAAGAGGTAGAGAGATGAGATTTTATTTGGAAGATTCATCTTA	900
OY	901	AGAGATGGACAGAGCTTTGTTGGCAATGTGTGATCCAAAGTTCTAGCTTAAAAAAAAGAC	960
Db	901	AGAGATGGACAGAGCTTTGTTGGCAATGTGTGATCCAAAGTTCTAGCTTAAAAAAAAGAC	960
OY	961	GACAAAGGAATGCGAAGAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGACATGAG	1020
Db	961	GACAAAGGAATGCGAAGAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGACATGAG	1020
OY	1021	GCTTTAGAAAGTTTATGTAAAGAAAATGGTTTAAAGTAATGATGAGACGAAAAAATGTGAA	1080
Db	1021	GCTTTAGAAAGTTTATGTAAAGAAAATGGTTTAAAGTAATGATGAGACGAAAAAATGTGAA	1080
OY	1081	GAAATGCAAAATGATTAATTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCATTAATAT	1140
Db	1081	GAAATGCAAAATGATTAATTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCATTAATAT	1140
OY	1141	CGTCTTTTGTATCCAAACAAAGAAATATGAAATTTGTTGATGAGGAAGGTTGCCAACA	1200
Db	1141	CGTCTTTTGTATCCAAACAAAGAAATATGAAATTTGTTGATGAGGAAGGTTGCCAACA	1200
OY	1201	TTTCTTACCAACGAAGATTTGTGCGAAATTTGAGATCTTATTTCTATTTTGAAAAAAAA	1260
Db	1201	TTTCTTACCAACGAAGATTTGTGCGAAATTTGAGATCTTATTTCTATTTTGAAAAAAAA	1260
OY	1261	TGTCAGATGGAAGAAATGCAATGTATTAATTAAGAGCAATGTATCAAAAGAGACTT	1320
Db	1261	TGTCAGATGGAAGAAATGCAATGTATTAATTAAGAGCAATGTATCAAAAGAGACTT	1320
OY	1321	GATGACAGGGCAAAATTAAGTGTGTCGAAGAAATATGCGAGGAATGTTACATGTTCAAC	1380
Db	1321	GATGACAGGGCAAAATTAAGTGTGTCGAAGAAATATGCGAGGAATGTTACATGTTCAAC	1380
OY	1381	AAACCTGCTGTTGAAGTTTCAACAAAGATTAATGTAATGATGTGAGAACTGAAAAAA	1440

Db	1381	AAAAAGCTGGCTTGAAGAAAGTTTCAACAGAAATTAGTAAAGTATGTGAAATCTGAAAAA	1440
QY	1441	GAAGAACAAAGAGCTTCTCAACGATGAATTATTTATCTGTGTGACAGCCAGCAAAA	1500
Db	1441	GAAGAACAAAGAGCTTCTCAACGATGAATTATTTATCTGTGTGTACAGCCAGCAAAA	1500
QY	1501	GCAGCCGGGTGCTTACACATGATCTTCGAATGAAAACTATCTTTTATACACACACATG	1560
Db	1501	GCAGCCGGGTGCTTACACATGATCTTCGAATGAAAACTATCTTTTATACACACACATG	1560
QY	1561	GATCAAAAGGAGATTCCCGACAGATAAAAATTGCAGGAAATTGGGAGAAAGTGCCAA	1620
Db	1561	GATCAAAAGGAGATTCCCGACAGATAAAAATTGCAGGAAATTGGGAGAAAGTGCCAA	1620
QY	1621	GATTTAGAGAGATTCCAAAGAAATTCATGCGCATGTGATCACTGAGACAGCAATGC	1680
Db	1621	GATTTAGAGAGAGATTCCAAAGAAATTCATGCGCATGTGATCACTGAGACAGCAATGC	1680
QY	1681	AATCGCTTGGGAGCTACAGAAATTTTAAACAGGTTTATTTGATGTGACCAAGATACT	1740
Db	1681	AATCGCTTGGGAGCTACAGAAATTTTAAACAGGTTTATTTGATGTGACCAAGATACT	1740
QY	1741	TTGAAAGACCAAGAACTGTGTAAATTAACCTAATAAAGTATTAATAGGTCTAGA	1800
Db	1741	TTGAAAGACCAAGAAAGTTGTAAATTAACCTAATAAAGTATTAATAGGTCTAGA	1800
QY	1801	AGAGGAGATGACCGTTCTCTTTTGTATGTGTCTCCAAAGCCGTACGTGTGAGTATG	1860
Db	1801	AGAGGAGATGACCGTTCTCTTTTGTATGTGTCTCCAAAGCCGTACGTGTGAGTATG	1860
QY	1861	GTAAAGACGTGAAAGACAGGTGTGAAGTATCCAAAAAAATATATAAGCTTCATATATT	1920
Db	1861	GTAAAGACGTGAAAGACAGGTGTGAAGTATCCAAAAAAATATATAAGCTTCATATATT	1920
QY	1921	ATTGAATTTCTTGAATAATATACAAATATAATACACACTGGAAAGAAATTTGGCCCTC	1980
Db	1921	ATTGAATTTCTTGAATAATATACAAATATAATACACACTGGAAAGAAATTTGGCCCTC	1980
QY	1981	TGCGATACGTATTGTSCAATAGATTTTCACTTAATGTGCACAGTCTTACGAAGAGAAATAGT	2040
Db	1981	TGCGATACGTATTGTSCAATAGATTTTCACTTAATGTGTCCAGTCTTACGAAGAGAAATAGT	2040
QY	2041	TGTACAAAAATCAAGAAAGCATCGGACCGGCTCATATAAGAAAGCCCTTGAAAGATGCT	2100
Db	2041	TGTACAAAAATCAAGAAAGCATCGGACCGGCTCATATAAGAAAGCCCTTGAAAGATGCT	2100
QY	2101	CTCAAAAGTAGAGCTTCAAGGAAATTTGACATGATTAATCTAAATCTGAACTCGCATTA	2160
Db	2101	CTCAAAAGTAGAGCTTCAAGGAAATTTGACATGATTAATCTAAATCTGAACTCGCATTA	2160
QY	2161	AGATATTGTACAGTAGGCGGAAACGTAAATTAATCCGCAATCAGTGACTTATGCAAAAGCT	2220
Db	2161	AGATATTGTACAGTAGGCGGAAACGTAAATTAATCCGCAATCAGTGACTTATGCAAAAGCT	2220
QY	2221	AACAACCAAGATTAATCTGTGAAAGAGTGATGAGGATCTTGAAGAGAACTCTGTGAGAAA	2280
Db	2221	AACAACCAAGATTAATCTGTGAAAGAGTGATGAGGATCTTGAAGAGAACTCTGTGAGAAA	2280
QY	2281	TTAGTGAAGAGATGTGGAAGAACAGTGCAAAGCATTTACCAACAGAAATTGACAAACCGCA	2340
Db	2281	TTAGTGAAGAGATGTGGAAGAACAGTGCMAAGCATTTACCAACAGAAATTGACAAACCGCA	2340
QY	2341	GCTGATCTAAAAAAAGATTATTAAGACATTAAGAGAACTTAAAGAAAGCTGCAGAGGAAGCA	2400
Db	2341	GCTGATCTAAAAAAAGATTATTAAGACATTAAGAGAACTTAAAGAAAGCTGCAGAGGAAGCA	2400
QY	2401	ATGAACAAGTCCAGTCTGTGTTTTGTCACTCATTTAAGAAAAAGAAATTAATGTATCAAAA	2460
Db	2401	ATGAACAAGTCCAGTCTGTGTTTTGTCACTCATTTAAGAAAAAGAAATTAATGTATCAAAA	2460
QY	2461	AGTAATAGCAAAAACAAGATTAAGATGCCGTTCAACGAGCTTCAAGATACCAAGAAAA	2520



Db 2461 AGTAATAGCAAAAAAAGATAGATAGATCCGTTTCAACGAGCTTCAAGATACCAACAAA 2520  
Qy 2521 CATGTGAAATATACACGAGGGGAGTTAAGGATGATCCGTAACAGAAATTAAGAACTAAA 2580  
Db 2521 CATGTGAAATATACACGAGGGGAGTTAAGGATGATCCGTAACAGAAATTAAGAACTAAA 2580  
Qy 2581 GCAATTTGATTTGGCAGCAGAAAGTATTTGAAAGATATAGATTTGAAAGAAAGATTTAT 2640  
Db 2581 GCAATTTGATTTGGCAGCAGAAAGTATTTGAAAGATATAGATTTGAAAGAAAGATTTAT 2640  
Qy 2641 AAATTTGAAATCAGATTGTCAGAAATTAAGAGGATTTGCAAAAGATTAAGAAAGTATGCAA 2700  
Db 2641 AAATTTGAAATCAGATTGTCAGAAATTAAGAGGATTTGCAAAAGATTAAGAAAGTATGCAA 2700  
Qy 2701 AAGATTAATAAGGCTTGTCCGAATCTGAAGCTCTGAGAGGTAAGCCGACGAAACAGTG 2760  
Db 2701 AAGATTAATAAGGCTTGTCCGAATCTGAAGCTCTGAGAGGTAAGCCGACGAAACAGTG 2760  
Qy 2761 ACAGAAAGTACAAAGCACTACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2820  
Db 2761 ACAGAAAGTACAAAGCACTACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2820  
Qy 2821 GAATGCAAAATCTTTACAGACAAACAGACATGGGTTACACAGACATGACACACAAGC 2880  
Db 2821 GAATGCAAAATCTTTACAGACAAACAGACATGGGTTACACAGACATGACACACAAGC 2880  
Qy 2881 ACGCTTCTATCATCTACATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2940  
Db 2881 ACGCTTCTATCATCTACATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2940  
Qy 2941 AAACCAACCAAGTACAGACAG 3000  
Db 2941 AAACCAACCAAGTACAGACAG 3000  
Qy 3001 GGGCTGAGATGAGTGGGTGAGATGTGATGAGGGGGGATGATGATGATGATGATGATGAT 3060  
Db 3001 GGGCTGAGATGAGTGGGTGAGATGTGATGAGGGGGGATGATGATGATGATGATGATGAT 3060  
Qy 3061 TTCTATGATTTAG 3072  
Db 3061 TTCTATGATTTAG 3072

## RESULT 2

US-09-762-724-11  
; Sequence 11, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3054  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. E. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(3054)  
US-09-762-724-11

Query Match 75.5%; Score 2318.8; DB 4; Length 3054;  
Best Local Similarity 86.0%; Pred. No. 0;  
Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;

Qy 4 GCGCGGCGGTCAAGCGCGCAGG---CAGCAGGACACAGAAATGATGATGAGAAATCAT 60  
Db 1 GCGCGGCGGTCAAGCGCGCAGGACGATACAGAGCATACAGGCAATATGATGATGAGAAATCAT 60  
Qy 61 GTTTTACCTTTAATTTCTAAAGAAAGATGACTAAGTACAGAGAAATGCAAAAAAACTA 120  
Db 61 ATTTTGGCGTTGATTTCTACAAAGAAATGCAATGCAATCATCAAAATGCAAAAAAGTTTA 120  
Qy 121 AAAAAATATTGGCCAGAAATGACTGAAGCAAACTAAATATTAAGAAATGACAGAAAA 180  
Db 121 GAAAAATATCTCGAAGGTTGAAAAAGCAATCACTAAGCATGAAAAAGTACATTAATG 180  
Qy 181 CTTAAAGTTTTCGAGAGATGAAAAAGCAGATACAAATATGCAAAAGATGAAAGCCAAAT 240  
Db 181 CTTAAAGATTTCTGTGGAATATGGGAAAGCAAGTAAACCAATATCAAAATGCAAGGCTTA 240  
Qy 241 ATTGAGAAAAATGACTACATCAAAAGAAACTTAAAGCAATTAATAAAAAAATTT 300  
Db 241 CAAGCCAAAGTTACGGGAAATGTACAAATTTTAAACAAAAAGCTAAGACCGCGTTA 300  
Qy 301 CAGATTAACGATTAAGATTAAGATGATTCATCTTTGAAATTAAGTCTATCAAAAG 420  
Db 301 ACAATTCATCAGATTAATTTGCAAAAGAGATGAAACGAAATGCTATTTTGGAGGA 360  
Qy 361 GATATTTCAAAAGCACTTAAGATGATTCATCTTTGAAATTAAGTCTATCAAAAG 420  
Db 361 GCATG---CATATATCTTGTAAGAAATTTGTAACAACTAAGAAATCTATGTTACAGAAA 417  
Qy 421 AAACGTATTAAGTTGCGGAGAAAGTCTTTTAAAGACACTTGTAGCATCTTAATGGA 480  
Db 418 AAACGTATTAAGTTGCGGAGAAAGTCTTTTAAAGACACTTGTAGCATCTTAATGGA 477  
Qy 481 TCAGTATATGTAAGAAAAAACTTAAGATTTGGCTGTCAATGGAGGGAAGTAT 540  
Db 478 ACAGAAACCAATGAAAAAACTGAAGAAATTTGCCAAGCTTGCAAGGGAAGTAT 537  
Qy 541 GAGTTTAAACAACTGTGTCTGAAACAGAAAGACATGTAAGATATTTTAATGAAAAA 600  
Db 538 GAATTTAAGGACTTGTCTTTGAACCAAAAAAGCGCGGAAATTTATTAAGAAAAA 597  
Qy 601 GATTAAGATGCGGACTCTTTAAACAGATGTTTACAGACACTAGAAAGTTTAAAAA 660  
Db 598 GATTAAGATGCGGACTCTTTAAACAGATGTTTACAGACACTAGAAAGTTTAAAAA 657  
Qy 661 GAAACATGCTGAAATTAACCTGAAACAAATGCTATTTTAACTGGAAATTTGGAGAGAC 720  
Db 658 GAAATATGCTGTAATTAACCTGAAACAAATGCTATTTTAACTGGAAATTTGGAGAGAC 717  
Qy 721 GATTAATTAATGATTAATGATTTGAGAGAAATGCAAGAACAAATATTTGCTTATATG 780  
Db 718 GATTAATTAATGATTAATGATTTGAGAGAAATGCAAGAACAAATATTTGCTTATATG 777  
Qy 781 CCACCAAGACCCGATTTTGAATCCAACTAAGGCCAGAGCTTCAATAGACAGAGATATAGG 840  
Db 778 CCACCAAGACCCGATTTTGAATCCAACTAAGGCCAGAGCTTCAATAGACAGAGATATAGG 837  
Qy 841 CTGGAAGATTTTAAAGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 838 CTGGAAGATTTTAAAGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
Qy 901 AGAGATGCGACAGCTTGTGTCATTTGTC---ATCAAGATTTAGTCTTTAAAAAATA 957  
Db 898 AGAGATGCGACAGCTTGTGTCATTTGTC---ATCAAGATTTAGTCTTTAAAAAATA 957  
Qy 958 GACGACAAAGAGAAATGCGAAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAACAT 1017  
Db 958 AATATCGAGAAATCAATTAAGATTTGCAAGATTAAGCAAAAATCTCATGAACAT 1017  
Qy 1018 GAGGCTTTAAGAAATTAATGAAGAAATGCTTAAGTATGATGAAGAAAGAAATGAT 1077  
Db 1018 GAACCTCTTAAAAATTTATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1077  
Qy 1078 GAAGAAATTTGCAAAATGATTAATTAACAAAATTTTCACTTCAAAAAGTCACTAAT 1137

Db 1078 AAGAAATTAGAGAAAGATTTAAAAAATGACCAAACTCTAAACCAACGATCTTTAAA 1137  
Oy 1138 AATCCTCTTTTGTATCTCAACAAAGAAATATGAAATTTGTGATGGGAAGGTTGGCA 1197  
Db 1138 AACCATCTTTATGATCC-----AATGATTAATTTGTGATGAGAAAACTGGCA 1188  
Oy 1198 ACATTTCTTAGACAAGAAATTTGCGAAATTTGAGTCCATTTGTTTCTATTTTGAATA 1257  
Db 1189 ACATTTCTTAACTAAAGAAAGTGTGCAAGATTTGGAATCTTATTTGTTTACTACAAAAA 1248  
Oy 1258 AAATGTCAGATGAGAAATGATGTAAATAATTAAGGCAACATGTTACAAAGAGGA 1317  
Db 1249 GCTTGTCCAAATGCCAAAGAAAGTGTATGAAATTTGAGGGCAGCGTTATTAAGAGAG 1308  
Oy 1318 CTGTATGACGGGCAAAATTAAGTGTGCAAGAAATATGCGAGATGTTCATGTTTCA 1377  
Db 1309 CTGTATGACGGGCAAAATTAAGTGTGCAAGAAATATGCGGTTATTAAGTGTTC 1368  
Oy 1378 AACAAAGCTGCTTGAAGAAATTTCAACAGATTTAATTAATGTAAGAACTGAAA 1437  
Db 1369 AATCAAGTTTGTCTTAAGAGTTTCAACAAAGATTTAAGTAAGTAAGAGCT--A 1425  
Oy 1438 AAGAAACCAAGGAAGTTTCTCAACGATGAATTAATTTCTGTGTGTAAGCCAGCA 1497  
Db 1426 AAGAAATTAAGGAAGTTTCCCAACGATGAATTTGTGTGTGTACAGCCAGCA 1485  
Oy 1498 AAGAGCCCGGTTGCTTACACATGATCTTCGATGAAATCTATCTTTTTCACACAA 1557  
Db 1486 AAGCTGACAGATTTACTTACACAGATCATCAATGAGGTTATCTTTTACGACACAA 1545  
Oy 1558 CTGATCAAAAGCGAATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGGAAGTGC 1617  
Db 1546 TTGGATCAAAAGAGATTTTCCGACAGATTAAGCTGCAAGAAATTTGGGGAAGTGC 1605  
Oy 1618 CAAGATTTAGGAGAGATTTCAAAAGAAATTAATGAGCCATGTCATACCTGAGAGCA 1677  
Db 1606 CAAGATTTAGGAAAGATTTCAAAAGAAATTTACATGACCATGTCATACCTGAGAGCA 1665  
Oy 1678 TGCAATCGCTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACAAAGAT 1737  
Db 1666 TGCAATCGCTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACAAAGAT 1725  
Oy 1738 ACTTGAAGAAAGCAAGAAATTTGTGTAATTAATCTTAAAGAAAGTGTATTAATGTCCT 1797  
Db 1726 ACTTGAAGAAAGCAAGAAATTTGTGTAATTAATCTTAAAGAAAGTGTATTAATGTCCT 1785  
Oy 1798 AGAAGAGGATGACCGTTTCTTTGTATGTGCTTCCAAAGGCTACGTTGAGGCTG 1857  
Db 1786 AGAAGAGGATGACCGTTTCTTTGTATGTGCTTCCAAAGGCTACGTTGAGGCTG 1845  
Oy 1858 ATGCTAAAGACGTGAAGACAGTGTGTAATTTCAAAAAAATATTAAGCTTCATAT 1917  
Db 1846 ATGCTAAAGACGTGAAGACAGTGTGTAATTTCAAAAAAATATTAAGCTTCATAT 1905  
Oy 1918 ATTATTAATTTCTTGAATAATTAACAATAATTAACACACTGGAAGAAATTTGTCCT 1977  
Db 1906 ATTATTAATTTCTTGAATAATTAACAATAATTAACACACTGGAAGAAATTTGTCCT 1965  
Oy 1978 TCTTGCAATAGTATTTGCAATGATTTTCACTTAATTTGCGAGCTTACGAAAGAGAT 2037  
Db 1966 TCTTGCAATAGTATTTGCAATGATTTTCACTTAATTTGCGAGCTTACGAAAGAGAT 2025  
Oy 2038 AGTTGTACAAAAATCAAGAAAGCATCGTGACCGTTCTATTAAGAAAGGCTTTGGAAGAT 2097  
Db 2026 AGTTGTACAAAAATCAAGAAAGCATTTGTGACCGTTCTATTAAGAAAGGCTTTGGAAGAT 2085  
Oy 2098 GCTTCAAAAGTGAAGCTTCAAGAAATTTGACTGATTAATCTTAATTTGAACCTGCATTTG 2157  
Db 2086 GCTTCAAAAGTGAAGCTTCAAGAAATTTGACTGATTAATCTTAATTTGAACCTGCATTTG 2145  
Oy 2158 AAAAGATATTTGTAAGTGAAGGCGGGAACGTAATAATTTGCTCAATGAGGCTTATGCAAA 2217

Db 2146 AAAAGATATTTGTAAGTGAAGGCGGGAACGTAATAATTTGCTCAATCAGTGCCTATGCAAA 2205  
Oy 2218 GCTTAACCAAGATTAATCTTGAAAGAGTATGAGATGCTAGAAAGAACTCTGTGAG 2277  
Db 2206 GCTTAACCAAGATTAATCTTGAAAGAGTATGAGATGCTAGAAAGAACTCTGTGAG 2265  
Oy 2278 AAATTAAGTGAAGAGTGAAGAAACAGTGCAAACATTAACCAAGAAATTTGGAACAACG 2337  
Db 2266 AAATTAAGTGAAGAGTGAAGAAACAGTGCAAACATTAACCAAGAAATTTGGAACAACG 2325  
Oy 2338 GCAGCTGATCTTAATAAAGATTTATTAAGCATATGAGAACTTAAGAAACGTGAGAGAA 2397  
Db 2326 GCAGCTGATTTTAATAAAGATTTATTAAGCATATGAGAACTTAAGAAACGTGAGAGAA 2385  
Oy 2398 GCAATGAACAAAGTCAGTCTTTTGTGCTACTTAAGAAACGAAAGTAATGTATCA 2457  
Db 2386 GCAATGAACAAAGTCAGTCTTTTGTGCTACTTAAGAAACGAAAGTAATGTATCA 2445  
Oy 2458 AAAAGTAAATGCAAAAAACAAAGATTAAGAAATGCTTTCAACGAGCTTCAAGATTAACA 2517  
Db 2446 AAAAGTAAATGCAAAAAACAAAGATTAAGAAATGCTTTCAACGAGCTTCAAGATTAACA 2505  
Oy 2518 AAACATGTGAATAATTAACGAGAGGAGTTTAAGATGATCCGTAAACGAATTAAGAACT 2577  
Db 2506 AAACATGTGAATAATTAACGAGAGGAGTTTAAGATGATCCGTAAACGAATTAAGAACT 2565  
Oy 2578 AAACATTTGATTTGGCAGCAAGATTTTGAAGATTAAGATTTGGAAGAAAGATGT 2637  
Db 2566 AAACATTTGATTTGGCAGCAAGATTTTGAAGATTAAGATTTGGAAGAAAGATGT 2625  
Oy 2638 AATAAATTTGAATGCAATTTGCAAGATTTAAGAGATTTGCAAGATTTGAAGATTTG 2697  
Db 2626 AATAAATTTGAATGCAATTTGCAAGATTTAAGAGATTTGCAAGATTTGAAGATTTG 2685  
Oy 2698 AAAAGATTAATTAAGCTTGTGCAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACA 2757  
Db 2686 AAAAGATTAATTAAGCTTGTGCAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACA 2745  
Oy 2758 GTGACAGAAATGACAGCACTTACCAACCAACCAACCGTTGCCGATCCGAAGCA 2817  
Db 2746 GTGACAGAAATGACAGCACTTACCAACCAACCAACCGTTGCCGATCCGAAGCA 2805  
Oy 2818 ACGGAATGCAATCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2877  
Db 2806 ACGGAATGCAATCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2865  
Oy 2878 AGCAGCTTACTTACATCTTACCATCAATCAAAATTAATTAATTAATTAATTAATTA 2937  
Db 2866 AGCAGCTTACTTACATCTTACCATCAATCAAAATTAATTAATTAATTAATTAATTA 2925  
Oy 2938 TGCAAAACCAACCAAGTGTACGACAGAGAGAGATGTGACAGAACGTAACCGAGT 2997  
Db 2926 TGCAAAACCAACCAAGTGTACGACAG-----GGATGATGACAGAAAGCTGAAGCAAGT 2979  
Oy 2998 GAGGGGCTGAGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGATGTAATGCTTATTT 3057  
Db 2980 GAGGGCTGAGAGTGAAGGGGTGGAATGTGATGAGGGGGGTGATGTAATGCTTATTT 3039  
Oy 3058 TCCTTCATGATTTAG 3072  
Db 3040 TCCTTCATGATTTAG 3054

RESULT 3  
US-09-762-724-7  
; Sequence 7, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724

CURRENT FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: PCT/US99/18750  
 PRIOR FILING DATE: 1999-08-17  
 PRIOR APPLICATION NUMBER: US 60/096,805  
 PRIOR FILING DATE: 1998-08-17  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 7  
 LENGTH: 3084  
 TYPE: DNA  
 ORGANISM: Pneumocystis carinii sp. f. hominis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(3084)  
 US-09-762-724-7

Query Match 73.7%; Score 2262.6; DB 4; Length 3084; ...  
 Best Local Similarity 84.7%; Pred. No. 0;  
 Matches 2640; Conservative 0; Mismatches 399; Indels 78; Gaps 6;

QY 1 ATGGCGGGGGGTCACAGCGGACAGGACACAGAAATGATGAGAAAT 60  
 Db 1 ATGGCGGGGGGTCACAGCGGACAGGACAGAAATGATGAGAAAT 60  
 QY 61 GTTTACCTTTAATTCTAAGAGATGCACTAAGTAGCAGAAATGCAAAAACTA 120  
 Db 61 GTTTACCTTTAATTCTAAGAGATGCACTAAGTAGCAGAAATGCAAAAACTA 120  
 QY 121 AAAAAATATGCGCAAGATTGCTGAAGCAAACTAATATATGACACAGAAA 180  
 Db 121 GAAGATATGCGCAAACTATACAAATGCAAGATTAAATCCAGAAAAAGTTCAAGAAA 180  
 QY 181 CTTAAGGTTTTCGCAAGATGCAAAAGCAGATCAAAATGCAAAAGACTGAAGCCAA 240  
 Db 181 TTTAAAGATTTCTGTGATACGCGGAAACGAATGAAATGTCAGATCTTAAAAACAA 240  
 QY 241 ATTGAGAAAAATGTACTACATCAAAAGGAACTTAAAGAGAAATTAAGAAAAAT 300  
 Db 241 GTCAATCAAAATGTACTTAAATTCAGAGAAATCTTCAACAGCTGTAGAAAAATTT 300  
 QY 301 CAGATTAACGATTAAGATGCAAAAGATGACAAACAACTATTTTGAAGGA 360  
 Db 301 TCAGATTAACAGATGAGATTGCAAAAGATGACAAACAACTATTTTGAAGGA 360  
 QY 361 GTATGTTCAAAAGAACTTAAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAA 420  
 Db 361 GCATGTTCAACAGAACTTAAAGATGATGCAATTAATTAAGAAATTAAGTCTATCAAA 420  
 QY 421 AAAGGTGATTAAGTGGGAGAGAGTTCTTTAAGAGCACTTGTAGCGATCTTAATGA 480  
 Db 421 GAACGGAACATGTGGGAGAGAGTTCTTTGAGGCGCTTGTGTGATCTCAATGA 480  
 QY 481 TCACTCATATGTAAGAAAACTTAAAGATTTGCGCTGTCAATGAGGAGGAAAGTAT 540  
 Db 481 ACAAGACATGTGAAAAAACTGAAGAGTTTGGCGAAATTAAGAAAGAAAGGAT 540  
 QY 541 GAGTTAAACAACTGTGTCTGAACAGAAAGACATGTAAGATTTTAATTTGA 600  
 Db 541 GAATTTAACGAGCTTGTCTTTATCAAAAAACAATGCTGATGCTTGTAAACAAAGGA 600  
 QY 601 GATTAAGAGTGGGATCTTAAACAGATTTTCAAGAGCACTAGAGAAAGTTTAA 657  
 Db 601 AAAAGTAATGTGATCTTGTAAAGAGAGTTGAAGAGCACTTAAGAGAAATGAATTTG 660  
 QY 658 AAAGAAATGTCTTGAATTAATGCAACATGCTAATTTTAATTTGAATTTGCGAG 715  
 Db 661 CGAAGAAAAATGTCTAATTAATTTGAAGATGTTACTTCAAGAGGAACTGTGAAGGA 720  
 QY 716 -----ACGAGATATTAATATGT 735  
 Db 721 GACAAATCAAGTCAATTAATTAATTAAGCTGCAAAAGATATGTAACAGAGGT 780  
 QY 736 ATGAATTTGGAGGAAATGCAAGAACTAATATATGTTATATGCCACGAGCCGAT 795

Db 781 GAGGAATTAACAGAAAAAGTGGAAAAAGAAATATGTTATATGATCCAGATCCGAT 840  
 QY 796 TTTGATCAACTAGAGCCAGAGGCTCAATAGCAGAGATATAGGCTGGAAGCTTTAT 855  
 Db 841 TTTGATCAACTAGAGCCAGAGGCTCAATAGCAGAGATATAGGCTGGAAGCTTTAT 900  
 QY 856 AGAAGGTAGAGAGATGAGATTTTATTAAGAAAGATCATCTAAGAGATGCGACACT 915  
 Db 901 AAGAGGCGAAGAGATGAGATTTTGTGGAAGAAACATGTATAGATGCAACAGCT 960  
 QY 916 TTGTGGCAATGTTGATCAAGATTTAGTCTTAAAAAAAAGACGACAAAGAAATGC 975  
 Db 961 TTGTGGCACTACT-----TCTTAAGAAACCTTAAGAAAGAAATGT 1005  
 QY 976 GAGAGCCCTTCAAAAAAGCTGCAAAATCTCATATGAACATGAGGCTTTAAGAAATTA 1035  
 Db 1006 ATAAAGCCCTTAAGAAAAAGCTGCAAAACCTCATATGAATGAGGCTTAAGAAATCTA 1065  
 QY 1036 TGTAAAGAAATGATTTAAGTAATGATGAACGAAAAATGTGAAGATTCAGAAATGAT 1095  
 Db 1066 TGTAAAGAAATTAACCAAGTATGATGAACGAAAAATGTGAATGATTAAGAAAAAGAT 1125  
 QY 1096 ATTAACAAACTTGCAGAAATTTTCACTTCAAAAGTCACTAATAATCGTTTGTATCCA 1155  
 Db 1126 GTTAACAAACTTGTACAGCTTACATCAACATCTTAAGAAACCGCTTACATTTCA 1185  
 QY 1156 ACAAAGCAATATATGAATTTGTTGATGAGGAGGTTGCCACATTTCTTACAGAA 1215  
 Db 1186 CC-----TGAATGAATTTGCGGAAATGGGAAATTAACCGACATTTCTTAAGTGA 1236  
 QY 1216 GATTGCGAAATTTGAGTCTATTTGTTCTAATTTGAAAAAAAATGCCAGATGAGAA 1275  
 Db 1237 GATTGCGAAATTTGAGTCTATTTGTTCTAATTTGAAAAAAAATGCCAGATGAGAA 1296  
 QY 1276 AATCATGTAAATATATAGAGCAACATGTTACAAAAGAGACTTGTATGACGCGCAAT 1335  
 Db 1297 GAAGCTGTATGATATGAGAGGCGAGCTGTTACAAAGAGAGGCTTGTATGACGCGCAAC 1356  
 QY 1336 AAGTGTGCAAGAAATATGCGAGAAATGTTTACATGCTTCAACAAAGCTGCTGAA 1395  
 Db 1357 AGTGTGTTGCAAAAAATATGCGGCTTATATGCTGTTCAATCAAGTTGCTTAAG 1416  
 QY 1396 AAGTTCAACAGAAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1455  
 Db 1417 GAGTTCAACAGAAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1473  
 QY 1456 TTCTCAACAGATTAATTAATTTCTGTGTATCAGCCAGCAAAAGCAGCGCTTCTT 1515  
 Db 1474 TTCCCAACAGATTAATTAATTTCTGTGTATCAGCCAGCAAAAGCAGCAATTAAT 1533  
 QY 1516 ACACATATCTTGAATGAAGAACTATCTTTTACGACACATGATCAAAAGCAGAT 1575  
 Db 1534 ACACACATATCAAAATGAGGTTAATCTTTTACGACACATGATCAAAAGCAGAT 1593  
 QY 1576 TTCCCGACAGTAATTAATTTGCAAGAAATTTGGGAGAAAGTGCAGAAATTTAAGAGAGAT 1635  
 Db 1594 TTTCGACAGTAATTAATTTGCAAGAAATTTGGGAGAAATTTAAGAGAGAT 1653  
 QY 1636 TCAAGAAATTAATCATGAGCATGTCAATCACTGAGAGCAATGCAATCGCTTGGGACT 1695  
 Db 1654 TCAAGAAATTAATCATGAGCATGTCAATCACTGAGAGCAATGCAATCGCTTGGGACT 1713  
 QY 1696 ACAGAAATTTAAGCAGGTTTATTTGATGAACACAAAGATTTTGAAGAACCAAGAA 1755  
 Db 1714 ACAGAAATTTAAGCAGGTTTATTTGATGAACACAAAGATTTTGAAGAACCAAGAA 1773  
 QY 1756 AGTGTGTAATTAATCTTAAGAAAGAAAGTGAATTAAGTCTGAAGAGAGATGAGCGT 1815  
 Db 1774 AGTGTGTAATTAATCTTAAGAAAGAAAGTGAATTAAGTCTGAAGAGAGATGAGCGT 1833  
 QY 1816 TTCTCTTTGATGTTCTTCAAAAGCTTACGTGTGAGCTGAATGAAGACGTGA 1875



```
Db 481 ACAAGCATGTGAAAAAAGCTGAAAAGATTGCCGAAATTAAGAAAGAAAGCAT 540
Qy 541 GAGTTAACTAACTGTGTCTGAACCAAGAAAGCATGTAAAGATTTTATTTAGAAAA 600
Db 541 GAAATTAACGAGAGCTTTGCTTTTATCAAAAAACAACTGCGTAAGTCTTTGTAACAAAAGGA 600
Qy 601 GATTAAGAGTGCAGGTACTCTTAAACAGATGTTTACAGAGCACTAAGAAAGTTTAAA--- 657
Db 601 AAAAGTAAATGTGTAATCTTTGAAAAAGAAAGTTGAAGAGCACTTAAGAAAGATGAATTG 660
Qy 658 AAAGAAACATGCTCTTGAATTAATCGAAACAATGCTATTTTACATTGAAAAATGCGAG-- 715
Db 661 CGAAGAAAAATGTACTATTACTTGAAGCAATGTTACTTTTACAGAGGAACTGTGAAGGA. 720
Qy 716 -----ACGACGATATTAATTAATGT 735
Db 721 GACAAATCAAAAGTGCATTAATAACTTAATATTAAGACTGCAAAAGAAATATGTACAGAGTGT 780
Qy 736 ATTGAATTTGGAGGAAATGCGCAAGAAACAAATATTGCTTATATGCCACCGAGCCGAT 795
Db 781 GATGAATTAAGCAAAAAGTGTGAAAAGAAATATGTTATATGCAATCCAGATCCGAT 840
Qy 796 TTTGATCCAACTAGGCCAGAGGCTACAATAGCAGAGATATAGGGCTGGAAGATTTAT 855
Db 841 TTGATCCAACTAAGCCAGAGCTACACTAGCAGAGGACATAGGGCTGGAAGAGCTTTAT 900
Qy 856 AAGAAAGTAAAGAGATGAGATTTTATTTGAAAAGAAATCATCTTAAGAGATGCACAGCT 915
Db 901 AAGGGGACAGAAAGATGGAATTTTGTGGAAGACAAACATGTAAAGATGCAACAGCT 960
Qy 916 TTGTGGCATTTGTGATCCAAAGATTTAGTCTTAAAAAAAAGACGACAAAGAGAAATGC 975
Db 961 TTGTGGCACTACT-----TCTTAAGAAACCTTTAAAAAGAAAGATGT 1005
Qy 976 GAAGAAACCCCTTCAAAAAAGCTGCAAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTA 1035
Db 1006 ATAAAGCCCTTAAAAAAAGCTGCAAAACCTCATGAACATGAGGCTTTAGAAATCTA 1065
Qy 1036 TGTAAAGAAAAATGCTTTAAGTAATGATGAAACGAAAAAATGTGAAGATTCGAAATGAT 1095
Db 1066 TGTAAAGAAAAATAAACCAAGATGATGAAACGAAAAAATGTGATGAACCTAGAAAGAAAT 1125
Qy 1096 ATTAAAGAAATCTTCAAAATTTTCACTTCAAAAGCACTAATATGCTTTTGAATCCA 1155
Db 1126 GTTAAACAAACTTGTACAGTCTTACATCAACAATCTTAAAAACGCTTTTACATTTCA 1185
Qy 1156 ACAAAAGAAATATGAAATTTGATGGGAAAGGGTTGCCAATTTCTTAGCAAGAA 1215
Db 1186 CC-----TGAATGAATGGCGAATGGGGAATTTACCGACATTTCTTAAGTAGAA 1236
Qy 1216 GATTGTGCAAAATTTGAGTCTTATTTGTTCTATTTTGAAAAAAATGTCAGATGAGAA 1275
Db 1237 GATTGTGCAAAACTTAAGATCTTATTTGCTTTATTAAGAAACCTGTCAGATGTCAAA 1296
Qy 1276 AATGCAATGAATAATTAAGAGCAACATGTTACAAAAGAGGACTGTAAGCGGGCAAT 1335
Db 1297 GAAGCTGTATGAATGTGAGGGCAGCGTGTATTAAGAGGGGCTTGATGACCGGCAAC 1356
Qy 1336 AAAGTGTGCAGAAATATGCGAGAAATGTTACATGGTTTCAACAAAGCTGCTGAA 1395
Db 1357 AGTGTGTTCGAAAATAATGCGAGGGTATTTGATGCTCAAAATTAAGATTTGGCTTAAG 1416
Qy 1396 AAGTTTCAACAAAGATTAAGTAAGTAAGTATGTAGAAACGAAAAAAGAAACAAAGAAAT 1455
Db 1417 AAATTTCAACAAAGATTAAGTAAGTAAGTATGTAGAAACG---AAAGAAATTAAGAAAGT 1473
Qy 1456 TTCTCAACAGATGAATTAATTTATCTGTGTGTAAGCCAGCAAAAGACCGCGGTGCTT 1515
Db 1474 TTCTCAACAGATGAATTTGTTGTTCTGTGTATACCAACAGCAAGGACGACGATTACTT 1533
Qy 1516 ACACATGATCTTTCGAATGAAAACTATCTTTTACGACAAACAACTGATCAAAAGCGAGAT 1575

Db 1534 ACACATCACCATCAAAATGAGAGTTATCTTTTACGACAAACAACTGATCAAAAGAGAT 1593
Qy 1576 TTCGCCGACAGATTAATAATTTGCAAGAAATTTGGGGGAAAGTGCCAAAGATTTAGAGAGAT 1635
Db 1594 TTTCGACAGATTAAGACTGCAAGAAATTTAGGGAGAAATATGCCAAGATTTTAGAAAGAT 1653
Qy 1636 TCAAAAGAAATTCATGATGCCATGTCATACACTGAGAGCAATGCAATGCTTGGGACT 1695
Db 1654 TCAAAAGAAATTCATGATGCCATGTCATACACTGAGAGCAATGCAATGCTTGGGACT 1713
Qy 1696 ACAGAAATTTTAAAGCAGTTTATTTGATGGAACAAAGATTAATTTGAAAAGCAAGAA 1755
Db 1714 ACAGAAATTTTAAAGCAGTTTATTTGATGGAACAAAGATTAATTTGAAAAGCAAGAA 1773
Qy 1756 AGTTGTATAAATACCTAAAGAAAGAGTGTAAATTAATGCTTGAAGAGAGATGACCGT 1815
Db 1774 AACTGTCAAAATTAATTTAAAGAAAGATGCAATTAATGCTTGAAGAGAGATGATGCT 1833
Qy 1816 TTCTCTTTGTATGTGCTTTCCAAAACGCTACGTGTGAGCTGATGTAAAGAGCTGAA 1875
Db 1834 TTTTCTTTGTATGTGCTTTCCAAAACGCTACATGTGAGCTGATGTAAAGAGCTGCA 1893
Qy 1876 GACAGGTGAGATTAATCAAAAAAATTAAGAGCTTCAATATTAATTTGAATTTCTTGA 1935
Db 1894 GATAGGTGCAAAATTAATGAAAGAAATATGCAAGCATCAATATTAATGATTTCCCTTAA 1953
Qy 1936 AATTAATCAAAATTAATAACACCTGGAAGAAATTTGTCCTGCTTGCATACGATTTGC 1995
Db 1954 AAAATATCAAAATTAATAAGCAATACAGCAATATTTGTCCCTCATGATCCATACGTC 2013
Qy 1996 AATGATTTTACCTAATTTGTCAGGTCTTAACGAAAGAAATAGT---TGTACAAAAATC 2052
Db 2014 GATGATTTTACCCATTTGCTGATCTTAAGAAAGAAAACTTTCTGTCAAAATCTT 2073
Qy 2053 AAGAGCATGTGAGCCGCTTCTAATAAGAAAGGCCCTTGAAGATGCTTCAAGATGAG 2112
Db 2074 AAAAATATTTGCGAACCATTTCAAAAAAGAAAGTTTGAAGATGCTTAAAGTGAAG 2133
Qy 2113 CTTCAGAAATTAATGATGATTAATCTAAATGTGAACCTGATGAAGAAATATTGTACA 2172
Db 2134 CTTCAGAAATTAATGATTAATTAATCTAAATGTGAACCTGATGAAGAAATATTGTACA 2193
Qy 2173 GTAGCGGAAACGTAATTAATGCGTCAATCAGTGGCTTAATGCAAAAGCTTAACCAAGAT 2232
Db 2194 GTATTTGAAGACGTAATTAATGCGTCAATCAGCAGTTTATGTAAGTAATTAACGGAAGT 2253
Qy 2233 AACTTGAAG--AGTATGAGATGCTTAAGAAAGAACTGTGTGAATTTGTGA 2289
Db 2254 AAAAATTAAGAGCGATTAATAAATGTTAAGAAAGAGCTTTGTCTTAATAATTTGTGA 2313
Qy 2290 GAAGTGAAGAAACGTGCAAAAGCAATTACCAAGAAATTAGGACAAACGGGACGTGATCTA 2349
Db 2314 GAGGTGAAGACGATCAAAAGTATTAACCAACGAATTAACAGAGCTGGAAGAAAGTCTA 2373
Qy 2350 AAAAAGATTAAGACATATGAGAACTTAAGAAACGTGAGAGAAAGCAATGAACAAG 2409
Db 2374 AAAAAGATGTTAAGACATATGAGAACTTAAGAAAGGGCAAAAAGCAATGAACAAG 2433
Qy 2410 TCAGTCTTGTGTTGCTCACTCATTAAGAAAAAGAAAGTAATGTATCAAAAAGTAATGC 2469
Db 2434 TCAGTCTTGTGTTTATCACTTGTTAAGAAAAAGAAAGTAATGTATCAAAAAGTAATGC 2493
Qy 2470 AAAAAGAAATTAAGATGCGCTTCAAGGAACTTCAAGATTAACAAAAACATGTGAA 2529
Db 2494 AAAAAGAAATTAAGATGCGCTTCAAGGAACTTCAAGATTAACAAAAACATGTGAA 2553
Qy 2530 ATACTAGGAGGAGTTAAGATGTAATCGTAAACAGAAATTAGAAAGCTTAAGCAATTTGAT 2589
Db 2554 ATACTAGGAGGAGTTAAGAGGCACTGTGAACAAATCTGAAGCAAGCAATTTGAT 2613
Qy 2590 TTGGCAGCAGAGATTAATTTGAAGATATGTAATTTGAAGAAAGATGTAATTAATTTGAA 2649
Db 2614 TTGGCAGCAGAGATTTGGAAGATATGTAGACTTTGAAGAAAGATGTAATTTGAACT 2673
```

QY 2650 TCAGATTGCAGAAATTAAAGAGATTGCAAAAGCTTAGAAGATATGCAAAAAGATTAAAT 2709  
DB 2674 TCGGATTGCGGATTTAAAGACGATTGCGATTGTTAAAGAAAGTGTGAAAGATTGAG 2733  
QY 2710 AAGCTGTCCGCAATCTGAAGCTTCTGAGTGAAGCCGCAAGAAACGTGACGAAAGT 2769  
DB 2734 AAGACATGTCACGATCTGAAAGCTTCTGAGGTGAAGTGCATGAATGTCAAGAAAGC 2793  
QY 2770 ACAACGACAACTTAAACAACAACAACACCTGCGCATCCGAAAGGCAACGAAATGCAAA 2829  
DB 2794 ACAACGACGACCAACAACAACAACGCTTACCGATCCGAAAGGCAACGAAATGCAAA 2853  
QY 2830 TCCTTACAGACAAAGACACATGAGTTACACAGACATGACACACACAAAGCCTTACT 2889  
DB 2854 TCCTTACAGACAAAGATACATGAGTTACACAGACATGACACACACAAAGCCTTACC 2913  
QY 2890 ATCATCTTACCTTACATCTCAAAAAATACATTGACATCAACGAGCGGATGCAACCAAC 2949  
DB 2914 ATCATCTTACCTTACATCTCAAAAAATACATTGACATCAACGAGCGGATGCAACCAAC 2973  
QY 2950 AAGTTCACGACGAGAGAGAGATGATGACGAGACGTAACCGAGTGAAGGCGCTGAG 3009  
DB 2974 AAGTTCACGACGAG-----GAGTGAACGAGAGACGTAACCGAGTGAAGGAGATTGAG 3027  
QY 3010 ATGAGTGGGTGGAATGTGATGATGAGGGGGGTGATGATGAGCAATGTTATTCCTTCAATG 3069  
DB 3028 ATGAGTGGGTGAGCGGTGATGAGGGGGGTGATGATGAGCAATGTTATTCCTTCAATG 3087  
QY 3070 TAG 3072  
DB 3088 TAG 3090

## RESULT 5

US-09-762-724-9  
Sequence 9, Application US/09762724  
Patent No. 6664053  
GENERAL INFORMATION:  
APPLICANT: Kovacs, et al.  
TITLE OF INVENTION: Identification of a region of the major surface  
protein of *glycoprotein (MSG) gene of human Pneumocystis carinii*  
FILE REFERENCE: 4239-58054  
CURRENT APPLICATION NUMBER: US/09/762,724  
CURRENT FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: PCT/US99/18750  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: US 60/096,805  
PRIOR FILING DATE: 1998-08-17  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 3081  
TYPE: DNA  
ORGANISM: *Pneumocystis carinii* sp. f. *hominis*  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3030)  
US-09-762-724-9

Query Match 52.9%; Score 1624.8; DB 4; Length 3081;  
Best Local Similarity 72.8%; Pred. No. 0;  
Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

QY 1 ATGGCGCGGCGGTAGCGGAG-----GGAGAGGAGACACAGATAGCATTTGAT 51  
DB 1 ATGGCGCGGCGGTAGCGGAGCTGCAAAAGATCAAGGCGCTAGTGTATATGATGAT 60  
QY 52 GAGGAACATGTTTAACTTTAACTTAAAGAGATGAGCTAAGTGAAGAGATGCAAA 111  
DB 61 GAGGAATATCTTTGCTTTAATGCA---GGAAGAAATATTAATGATGAATGAATGCAAA 117  
QY 112 AAAAAAATTAATTTATGCGAAGATTGACTGAAGCAAAATTAATTAAGACAGTA 171

DB 118 AAAAAATTAAGAAATTTTAAAGACATTTAAGATGATGATGATTAAGAAACGAAAAAGTT 177  
QY 172 CACAGAAATCTTAAAGGTTTTTGGCAATGAGAAAAACAGATACAAATGCAAAAGAACTG 231  
DB 178 CACAAAAATCTTAAAGGTTTTTGGCAATGAGAAAAACAGATACAAATGCAAAAGAACTG 237  
QY 232 AAAGCAATATTTGAAAAAATGTAATCAATCAAGAGAAATCTTAAAGAACTTAA 291  
DB 238 AAAGAAAAATCTCACTCAAAATGTAATGCAATCAAGAGAAATCTTAAAGAACTTAA 297  
QY 292 AAAAAATTTGAGATTTAAGGATTAAGATTTGCAAGAGATGCAATGCTTATTT 351  
DB 298 AAAAAATTTGAGATTTAAGGATTAAGATTTGCAAGAGATGCAATGCTTATTT 357  
QY 352 TTGAGAGGAGATTTGCAAAAGAACTTAAAGATGATGCAATGCTTAAAGATGCT 411  
DB 358 TTGAGAGGAGATTTGCAAGGAACTTAAAGATGATGCAATGCTTAAAGATGCT 417  
QY 412 TATCAAAAGAAAGCTGATTAAGGTTGCGAAGAACTTTTAAAGACATTTGATGAT 471  
DB 418 TATCAAAAGAAAGCTGATTAAGGTTGCGAAGAACTTTTAAAGACATTTGATGAT 477  
QY 472 CTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531  
DB 478 CTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
QY 532 GAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591  
DB 538 GAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 597  
QY 592 ATTGAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 646  
DB 598 TTGAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 647 -GAGTTTAAAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705  
DB 658 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
QY 706 AATTGCGAG-----ACGACATTAATTT 729  
DB 718 AATTGCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777  
QY 730 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 789  
DB 778 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
QY 790 CCGATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849  
DB 838 TCGATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
QY 850 TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 909  
DB 898 CTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957  
QY 910 ACAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969  
DB 958 ACTGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1015  
QY 970 AATGCGAAGAGGCTTCAAAAGAGTCAAAAGATGATGATGATGATGATGATGATGATGATGAT 1029  
DB 1016 -AATGCGAAGAGGCTTCAAAAGAGTCAAAAGATGATGATGATGATGATGATGATGATGATGAT 1074  
QY 1030 AGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1089  
DB 1075 GATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134  
QY 1090 AATGATATTAACAAAGCTTCAAAAGATGATGATGATGATGATGATGATGATGATGATGAT 1149  
DB 1135 AAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194  
QY 1150 GATCAACAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1209



Db 1195 -----GGATCCGGAAGATCATTCATGATATAGTACGACATTTCTTAGT 1242  
 Qy 1210 AACGAAGTTGTGGAAATTGGAGCTATGTTTCTATTTTGAAGAAATGTCAGAT 1269  
 Db 1243 GACAAATGCTGCACAAAGTTAGAGTCAGACTGTTTATTTT- -AAAGTCAAGCAGCT 1299  
 Qy 1270 GGAGAAATGTCATGTAAATAATATAGACCAATGTTACAAAGAGCACTTGATGACCG 1329  
 Db 1300 CTTGACAAAGAAATGTAAATCTGAGGACAGATGTTATAGAGAGGGCTTGAGACAA 1359  
 Qy 1330 GCAATTAAGTCTGCAGAAATATGCGAGAAATGTTACATGTTCAACAAAGCTGG 1389  
 Db 1360 GCTAATGAGCATTCGAGAAAGATGACGACTGTTCTATGTTCCAGGCAAGAAATGG 1419  
 Qy 1390 CTTGAAAGTTTCAACAGAAATTGATAAGATGTGAAACCTGAAAGAAAGAAACAA 1449  
 Db 1420 TTTAAGAACTACTAGAAATATATGAGAAATGTTGGAATTTAAAC- - - - - 1470  
 Qy 1450 GGAAGTTTCTCAACGATGAAATTATTTCTGTGTGACAGCAGCAAGAGCCCG 1509  
 Db 1471 -----ACAAGCATGAGTGTGTTTGTATGATGATCCACTTAAAGCAGTCA 1521  
 Qy 1510 TTGCTTACATGATCTTCGAATGAAATCTATCTTTTGAAGACAACTGATCAAG 1569  
 Db 1522 ATACTTCAGCTGATATCCAAAGAGCAGCTTTTTCGGAACCAATGATCAAAAG 1581  
 Qy 1570 CGAGATTTCCGAGATTAATAATTCAGAGAAATGGGAGAAAGTGCAGATTTTGA 1629  
 Db 1582 CGAGATTTCCAGAGAAAGATTCAGAGAAATTTGAGAAAGATGGAAGCTTTAGGG 1641  
 Qy 1630 GAGATTTCAAAAGAAATTAACATGCGCATGTCATACATGAGACAGCAATGCGTTG 1689  
 Db 1642 AAGGATTCATATCAGATTAAGTGCATGTCATACGCTTAAACAAAGTGTGATCGCTG 1701  
 Qy 1690 GGGACTACAGAAATTTTAAAGCAGGTTTATGATGACCAAGAAATGATCTTTAAAGAC 1749  
 Db 1702 GGGACTACAGAAATTTTAAAGCAGGTTTATGATGACCAAGAAATGATCTTTAAAGAC 1761  
 Qy 1750 CAAGAAAGTGTGTAATAATTAAGAAAGAAAGTGTATTAAGTCTAGAGAGGAGAT 1809  
 Db 1762 CATAAAACTGTACGAATATTTAAAGAAAGTGTATTAAGTCTAGAGAGGAGAT 1821  
 Qy 1810 GACGTTTCTCTTGTATGTCCTCCAAAGCCTACGTGAGATGATGTTAAAGAC 1869  
 Db 1822 GATGTTTCTCTTGTATGTCCTCCAAAGCCTACGTGAGATGATGTTAAAGAC 1881  
 Qy 1870 GTGAAGACAGGTGTGAATTAACAAAGAAATTAAGAAAGTGTATTAAGATTT 1929  
 Db 1882 GTGAAGACAGGTGTGAATTAACAAAGAAATTAAGAAAGTGTATTAAGATTT 1941  
 Qy 1930 CTTGAAATATATACAAATTAATAACAACTGAGAAAGAAATTTGCTTGGCATACG 1989  
 Db 1942 CTTGAAATATATACAAATTAATAACAACTGAGAAAGAAATTTGCTTGGCATACG 2001  
 Qy 1990 TATGCAATAGATTTTCACTAATTTGTCAGAGTCTTACGAAAGAAATAGT- - -TGTACA 2046  
 Db 2002 TATGCGATAGATTTTATCCCAATTTGTCATCTTAAAGAAAGAAATTTTGTCTCA 2061  
 Qy 2047 AAAATCAAGAGCATGTCAGCGTCTTATTAAGAAAGAAAGTGTGAGATGCTCTCAA 2106  
 Db 2062 AATCTTAAAAATATTTGGCAACATCTTACAAAGAAAGTGTGAGATGCTCTTAA 2121  
 Qy 2107 GTAGAGTTCAAGAAATTTGATGATTAATCTAAATGTGAACCTGCTTGAAGATAT 2166  
 Db 2122 GTAGAGTTCAAGAAATTTGATGATTAATCTAAATGTGAACCTGCTTGAAGATAT 2181  
 Qy 2167 TGTACAGTACGAGAAAGTAAATTAATGTCGCAATCACTGATGCTTATGCAAGTCAAC 2226  
 Db 2182 TGTACATATTTGAAAGAAATGTAAGATTCATCAATCAACAGTTTATGTAAGATTAATCC 2241  
 Qy 2227 AAGGATTAATCTTGAAAGATGATGAGATCTTGAAGAAAGTCTGTGAGAAATTAAGT 2286  
 Db 2242 GAAAGTAAATCTAAAGAAAGCAGATTAATGAAAGTGAAGAGCTTGTCTAAATTAAGT 2301

Qy 2287 AAAGAGTGAAGAACAGTCCAAAGCATTTACACAGAAATTAGCAACCGGACGCTGAT 2346  
 Db 2302 GAAGAGTGAAGAACAGCATTTAAATGTTTACAGCAAAATTTGAGATGAGAAAGAAC 2361  
 Qy 2347 CTAAAAAAGATTTAAGACATATGAGAACTTAAGAAAGTGTGAGAAAGCAATGAC 2406  
 Db 2362 CTAAAAAAGATTTTGAACATTTGAAAGAACTTAAAAAAGCGCGAGAAAGCAATGAT 2421  
 Qy 2407 AAGTCCAGTCTGTTTGTGATCATTAAGAAAAAGAAAGTATGATCAAAAGTAT 2466  
 Db 2422 AATCCATCTGTTTATATATCTGTTAAGAAAGATTAATTAATCAATGAAAGTAT 2481  
 Qy 2467 AGCAAAAAAGATTAAGAAATGCGTTTCAACGCACTTAAGATTAACACAAACATG 2526  
 Db 2482 AGCAAAAGCAAGATTAAGAAATCCGTTTCAACGCACTTAAGATTAACACAAACATG 2541  
 Qy 2527 AAAATACAGAGAGGAGTTAAGATGATCCGTAACAGATTAAGAAAGTAAAGCATTT 2586  
 Db 2542 AAAATACAGAGAGGAGTTAAGATGATCCGTAACAGATTAAGAAAGTAAAGCATTT 2601  
 Qy 2587 GATTGGCAGCAAGATTAATGAAAGTATGATGATTTGAAGAAAGATTAATTAATG 2646  
 Db 2602 GATTGGCAGCAAGATTAATGAAAGTATGATGATTTGAAGAAAGATTAATTAATG 2661  
 Qy 2647 GAATCAGATTCAGAAATTAAGAGATTTGCAAGACTTAAGAAAGATTAAGAAAGATTT 2706  
 Db 2662 GAATCAGATTCAGAAATTAAGAGATTTGCAAGACTTAAGAAAGATTAAGAAAGATTT 2721  
 Qy 2707 AATTAAGCTTGTGCAATCTGAAGCTTGTGAGAGTGAAGCGCAGAAACAGTACAGAA 2766  
 Db 2722 CAAGAGATTAATGTTGAATTAATAACCTGAAGATGAAGCGCAGAAACAGTACAGAA 2781  
 Qy 2767 AGTCAACGACAACTACAAACAAACAAACCGTTGCCATCCGAGCGCAACCGAATGC 2826  
 Db 2782 AGCAACAGACAGACAGACAGACAAACAGACCTGTATCTGATTCGAAAGCGCAACGAAATGC 2841  
 Qy 2827 AATCTTACAGACAAACAGACACATGCGTTTACACAGATTCGACACACAGACGCTCT 2886  
 Db 2842 AATCTTACAGACAAACAGACACATGCGTTTACACAGATTCGACACACAGACGCTCT 2901  
 Qy 2887 ACTATCAATCTTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2946  
 Db 2902 ACCATCAATCTTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2961  
 Qy 2947 ACCAAGTACAGACAGAGAGAGAGATGATGACAGAAACGTAACCGAGTGAAGGCGCTG 3006  
 Db 2962 ACCAAGTACAGACAGAGAGAGAGATGATGACAGAAACGTAACCGAGTGAAGGCGCTG 3015  
 Qy 3007 AGATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3066  
 Db 3016 AAGATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3075  
 Qy 3067 ATTTAG 3072  
 Db 3076 ATTTAG 3081

RESULT 6  
 US-09-762-724-1  
 ; Sequence 1, Application US/09762724  
 ; Patent No. 6664053  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kovacs, et al.  
 ; TITLE OF INVENTION: Identification of a region of the major surface  
 ; FILE REFERENCE: 4239-58054  
 ; CURRENT APPLICATION NUMBER: US/09/762, 724  
 ; CURRENT FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750  
 ; PRIOR FILING DATE: 1999-08-17  
 ; PRIOR APPLICATION NUMBER: US 60/096, 805  
 ; PRIOR FILING DATE: 1998-08-17



Qy 1964 AAGAAATTCCTCTTGGCATAGTATTCGATATTTTCACTAATTCCTCAGGTC 2023  
Db 1964 GACATATTTCTCTATTTGGCACCACCTGCCATTTTGGACCCCAATTCGCCGATC 2023  
Qy 2024 TTACGAAAGAAATAGTGTACAAATAATCAGAAAGCATCGTACCGCTTCATTAAGAA 2083  
Db 2024 TTGAAAAA-----AATTAATGTGAACATTTGAAAAAATTTGAAAACCTTATATATAGCAA 2080  
Qy 2084 AGGCTTGGAAAGATGCTCTCAAGTAGAGCTTCAAGAAAAATGACGTATTAATCTAAT 2143  
Db 2081 GAGACCTTGAATAAGCACTATATTTAGTTAGAGACATCTTGTAGAAAAAAACT 2140  
Qy 2144 GTGAACCTGCATTAAGAAATATTTGTACAGTGGGGAAGTAAATATTCGTCATCA 2203  
Db 2141 GCAAAACAAATCTTGATTAAGTACTGTACCTATGGAATCAACAGAAATTAACACTTA 2200  
Qy 2204 GTGGCTTATGCMAAGCTAACCAAGATTAATCTGAAAGATGATGAGGATCTAGAA 2263  
Db 2201 AAGGTTTTTGA-----ACAGTTCTAGTATTAACATGAAACATTTAGAG 2245  
Qy 2264 AGGAACTCTGTGAATAATTAAGTGAAGAGTGAAGAACAGTGCMAAGCATTAACAAG 2323  
Db 2246 ATAAACTTTGCAAAAACTAGTTCAGCGGTGAAGAAAAATGCAAGATTAACAAG 2305  
Qy 2324 AATTAGACAACCGCGAGCTGATCTAATAAAAAAGTTAAGACATATGAGGAATTAAGA 2383  
Db 2306 AACTTGAATAAGCAAAAAATGATTTAGAGAAAACTTAAGATTAAGAAAAATTAAGAA 2365  
Qy 2384 AACGTGCAAGAGAACATGAAACAGTCCAGTCTTGTTCATCTCATTAAGAAAAAG 2443  
Db 2366 AGGATACAAAAATGCAATGGAAGAAACAAATCTGTTTTTCAACAATTAATCAACAG 2425  
Qy 2444 AAGTATATGATCAAAAAATGATATGCAAAAAAAGGATTAAGATCCGCTTCAACGAGC 2503  
Db 2426 ATATATAACAGAAAAAG-----AGTCAACCTAGTACCGCTGTGTAG 2470  
Qy 2504 TTCAAGATCCACAAAAATGTAATAAAGCTTACGAGGAGGAGTTAAGAGTATCCGTA 2563  
Db 2471 TTCAAGATTTGTATCATTTT-----AAGCTTGAATAAGAAATGAAAAAGTTCAAGTGA 2524  
Qy 2564 CAGAAATTAAGAAAGTAAAGCATTTGATTTGGCAGCAGAAATTTTGAAGATATAGATT 2623  
Db 2525 CAGAAAAAGAAAGCAAAAGCGTTGATTTGGTAGCACTAGCATTCAGCTTTATGTAGAGT 2584  
Qy 2624 TGAAGAAAGATGTAATTAATTTGAAATCAGATTCAGAAATTAAGAGATTCAGAACT 2683  
Db 2585 TAAAAAGAAAGCTGTCAACATCTAAGAGATGATTCGAATTTAGAAAAAGAAATGTAAA--T 2641  
Qy 2684 TAGAAGAAATTAAGAAAAAGTTAATAAGCTTGTCCCAATCTGAAGCTCTGAGGTGA 2743  
Db 2642 GTAAAGACCAAGTGAAGATTAAGAAAAATATTTTAAATAATGAAACCATGTAAGTA 2701  
Qy 2744 AGCCGCAAGAAACAGTGCAGAAAGTACAAAGCAACTAACAAACAACAAACCGTTG 2803  
Db 2702 AGCCACATGAAATTAATAACAGTACGAAACCAATTAACAAGTGCAGAAACAGTCA 2761  
Qy 2804 CCGATCCGAGGCAACGG-----AATGCAATCTTACAGAACAAAGACA 2848  
Db 2762 AAGAAAGCAAAAAACAGTAGAGAGACGGAAGAAATGCAAAATCTCTGACCAACAGCA 2821  
Qy 2849 CATGGGTTCACACAACTCGACACACAAAGCAGTCACTATCACTTATCACTCAAT 2908  
Db 2822 CGTAGGTGCAAAAGCTCAACCAATACACACCTTCAAGACTGCTCAAGTAACTG 2881  
Qy 2909 CAAAAATTAACATTTGACATCAACGAGCGATGCAAAACCAACCAAGTGTATGCAAGAGAGG 2968  
Db 2882 CAABAATTAACCTGACCTGACGAGCGGTGTAAAGCTTAAGAAATGTATGCAAGAGAGG 2941  
Qy 2969 AAGATGATGCAAGAGAGTGAACCGAGTGGGGCTGAGATGATGTGGGTGGAATGTGA 3028  
Db 2942 AAGATGAAGCAGAGAGGTGAAGCCGAGTGAAGGGGCTGAAGATGATGGGTGAGTGTGA 3001

Qy 3029 TGAGGGGGGTAGTAGAGCAATGATTAATTCGTCATGATT 3069  
Db 3002 TGAGAGGGGTGTATTATTAAGCAATGATGATTCATTCATGATT 3042

RESULT 7  
US-09-762-724-3  
; Sequence 3, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 3006  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3006)  
; US-09-762-724-3

Query Match 30.2%; Score 927.4; DB 4; Length 3006;  
Best Local Similarity 60.3%; Pred. No. 2,2e-193;  
Matches 1859; Conservative 0; Mismatches 1131; Indels 93; Gaps 16;

Qy 2 TGCGCGGGGCGGTCAACGCGCAGCGCAGGACACAGAAATGCAATGATGAGGAACATG 61  
Db 2 TGCGCGGGGCGGTCAACGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAG 58  
Qy 62 TTTTACGTTTATTTCTTAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 121  
Db 59 TTTTGGCTTTGATTTTGAATAATGATATGAATAATGATATGAATAATGATATGAATAAT 118  
Qy 122 AAAAAATATTCGCAAGATTTGCTGAAGCAAAATTAATATTAAGCAAGTACACAGAAAC 181  
Db 119 AAGATATTTGTGAAGTTTAAAAATTTAAACAAAAAGAACCAAAAACTAGAAAGAAAGT 178  
Qy 182 TTAAGGTTTTTTCGAAAGATGAAAAAGC---AGATACAAAATGCAAGAACTGAAGGCA 238  
Db 179 TAGACGAATCTGCAAAAGATGATTAATAACAAATTAAGAACAAATGCAAGAAATCA 238  
Qy 239 ATATTTGAAAAAATGATCTACATCAAGAAAACTTAAAGAAAGCAATTTAAAAA 298  
Db 239 AGGTTAAAGCAAAAGTGTACTAGTTTCAAAACAGAACTGTAAAGCAGTCAAAAAAGGAG 298  
Qy 299 TTCAAGATTATTAAGGATTAAGGATTTGCAAGGAAATGAACAACTTGTGAGAAATTA 358  
Db 299 CTTCACATTTAAGAAATTAAGATTAAGATTTGTAAGAAATGAACAACTGCTGTTTGA 358  
Qy 359 GAGTATGTTCAAAAGAACTTAAAGATTTGCAATTAAGTATTTTGAAG 418  
Db 359 GAGATGTTCCAAAGAACTTAAAGATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 418  
Qy 419 AAAAAAGTGAATAGTTGCGGAAGAAAGTCTTTTAAAGCACTTCGTAAGCAATTTAATG 478  
Db 419 AAAAAAGTGAATAGTTGCGGAAGAAAGTCTTTTAAAGCACTTCGTAAGCAATTTAATG 478  
Qy 479 GATCAGTCATATGTAAGAAAAAATTAAGAGATTTGCTGTCATGGGAGGAAAGT 538  
Db 479 ATAAAAACATGCAAAATTAAGTAAAGGAGGTTTGTCAAGATTTCAACAAAGAAAGT 538  
Qy 539 ATGAGTTAACAACTTGTGTGTAACCAAGAAAGACATGTAAGATATTTTAAATGA 598

539 ATGACCTAATAAATTATGTCTTGACGAAGAAAAACGTGTGAGATCTGTATCTAAGA 598  
599 AAGATTAAGAGTCGCGTACTTTAAACAGATGTTTCAGACGACTAGAA-----GTT 652  
599 AAGAAATACAAATGCGAACTCTCAAAAGAAAGAAATGATCTAGTCTTGAAAGAAAGATT 658  
653 TTTAAAAAGAAACATGCTCTGAATTACTCGAAACATGCTATTTTAACTTGGAAATTCG 712  
659 TATTAAAAAGAAAAATTTATTATTCTTGAAAGAAATGTTACTTTATGCGTCAACCTGTG 718  
713 GAGACGACGATTAATAATTATGATTTGAATTGGAGAAAAATGCCAAGAACAAATATTG 772  
719 AAACGATCAGCA---AAGTCTAAAGATTGCAAGCAAAATGCTAAAGAAAAATCTCG 775  
773 CTATATGACCACGAGACCCGATTTTGATCCAACTAGGCCAGAGGCTCAATAGCAGAG 832  
776 TTTATGACGACGACGAGTTCACCTTTGATCTCCAGCAAAATTAAGATTAGTTAGCAGAG 835  
833 ATATAGGCGCTGGAAGATTTTATAGAAAGTGAAGAGATGAGATGTTTATTTGGAAGA 892  
836 AAATGACCTAGAAAAAATTTGACGTAGAAAGCAGTAAAAAGGAAATTCATTTGGAGGC 895  
893 ATCATCTAAGAGATCGACAGCTTTGTTGCACTGTTGATCCAGATTTCTAGCTTAAAA 952  
896 CATCAAATTAAGATTAAGTCCGCTTTATTTGCACTTATTAAGCA-----GAGTATGCTC 949  
953 AAAAAAGCAGCAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATG 1012  
950 AAAATACCTTTTAAAGATCAATGTGAAGATGTATTAAAAAAAATGTGAAAACCTTAAAG 1009  
1013 AAATGAGGCTTTAAGATTTATGTAAGAAAAATGTTTAAAGTAAAGTAAAGACGAA 1072  
1010 AGCATATTATTATTTAAAGATTTATGTAGTAAATGAACCTATCATCTATATCCAAAAAGAA 1069  
1073 AATGGAAGAAATTCGCAAAATGATATTAACAAAACCTTGCAAAATTTCACTTCAAAAGTCA 1132  
1070 AATGGAAGAACTAATAAAGAGGT-----AACAAACCGTATTTTAACTGTTCTAAA 1123  
1133 CTAAATATGCTCTTTTGAATCCACAAAGAAATGAATTAATGTAATGTTGATGGAAGGCT 1192  
1124 GGAATGAAATAATATTTCCCTCAGCTAAATGTAAA---GGAATATATGTTGGCATATGT 1180  
1193 TGCCAACTTTCTTGCAACGAAGATGTGCGAAATTTGGAAGCTATGTTCTTCTTTTG 1252  
1191 TGCATACATTTCTTGTGTAAGAGAGGTGACGAACTGTTGCGAATGTTTAAATTTGA 1240  
1253 AAAAAAAATGTCAGATGAGAAATGCAATGTAAAAATTAAGACCAACATGTTCAAAA 1312  
1241 AAAGCCAGCTCCT---GAAAAGCCCTGCAATTACTTAAGACAGATGTTTAA 1297  
1313 GAGACTTGATGACGCGCAAAATAAGTGTGCAAGAAATTAAGCAGAGATGTTACATG 1372  
1298 AAGGCTTGAAGCAATGACAAATGAAGCAATTACAAATTAAGTGAAGGGAATAATGCAAG 1357  
1373 GTTCAACAAAAGCTGGCTGAAAAGTTTCAACAAGATTGTAAGATGATGTAAG 1432  
1358 GTTCAAAATGAACATGCGCTGTAACCTTCAAAAAAATGTTAAAGTTGTGAAAA 1417  
1433 TGAATAAAAGAAAAACAAGAGATTCTCAACGATGAATTTATTTCTGTGTGACAGC 1492  
1418 CGAAAGGAGA-----AAGTGAATTAATTTGTACTATGTAAGAAC 1459  
1493 CAGCAAAAGACCCCGGTTGCTTACACATGATCTTGAAATGAAAACTATCTTTTACAGAC 1552  
1460 CAATTAATAACGCGCTTACAGTGTCAACAGATTTCCGAATGAGGACAGTGTGTTCAAG 1519  
1553 AACAATGGAATCAAAAGGAGATTTCCCGACAGATTAATAATGCAAGAAATGGGAGAA 1612  
1550 AGCATTTGAAGCAAAACGAGATTTTCAACAGAAAAGATGTAAAGATTAAGAAAA 1579  
1613 AGTGCCAAGATTTAGAGAGATTCAAAAGAAATTAACATGCGCATGTCACTACGAGAC 1672  
1580 AATGTGAGGCTTTAGGAAAAAGATTCAAGAGAAATTAATGTGTCACTTATACGTTAAAC 1639

1673 AGCATGCAATCGCTTGGGCACTACAGAAATTTTAAAGCAGTTTATTTGATGAACACA 1732  
1640 AGCATTCGAATCGGCTAGAGAGCATAGACCTTAGAAGAGAGATTTGCTAAAGAAAAATA 1699  
1733 AAGATCTTTGAAAGACCAAGAAAGTTGTGTAATATCTTAAAGAAAAAGTGAATAAT 1792  
1700 AAGATATTTTAAAGATGAATAATGCTGCAAAAGAAAGCTTAAGAAAGATGTGAAAAAT 1759  
1793 GGTCTAGAGAGAGATGACCGTTTCTTTTGTATGTCTTCAAAAACGCTACGTTG 1852  
1760 GGTTAGAAGAGAAATAATTAATTTTTCGCTGTTGTGACTGTGAACTTGTTGCA 1819  
1853 AGCTGATGCTAAAGACGTGAAGACAGTGTGAATTTCAAAAAAATTTAAAGCTT 1912  
1820 AAAAGATCACTAGAAATGTGAATCTAAATGTAATATATGAAGAGCATATGGAACCTA 1879  
1913 CATATATTATGAATTTCTGAAATATATCAATTAATAAATTAACAACATGGAAGAAAT 1972  
1880 TGAACGTTATTAAGTGAATTAAGCTTAAAAAGAGAAAAAATGTGAAAT-----TT 1930  
1973 GTCCCTTGGCATACGTAATTCGAATGATTTTCACCTAATGTCTCAGGCTTACGAAG 2032  
1991 GGGCTCCATATGTATAAAGATACAGCAAAATTTGTAAGAAACCTTAAAAACGAGAAAAAG 1990  
2033 AGAATAGTTGTACAAAATCAAGAGCATGTGACCGCTTATTAATAAGAAAGCCTTGG 2092  
1991 ATGGCAATGCAAAAACTCAATTAATAAGTCAATCATTTCTTGAAAAAAGAACCTTAA 2050  
2093 AAGATGCTCTCAAGTGAAGCTTCAAGAAAAATGACATGAATTAATCTTAATGTGAACCTG 2152  
2051 AAAATTAAGTTGTGAAGAAATTTGAAGGTATTTATCAACGTTAGAGAAATGTAACAATA 2110  
2153 CATTGAAGAATATTTGACAGTACGCGGAAACGTAAATATGCTCAATCAGTGGCTTAT 2212  
2111 CACTTAATATATCTGTACACAAATTTGAAGAAAGCAGAGATGGGTGGAACCTTGT--- 2167  
2213 GCAAAAGCTAACCAAGATTAATCTTGAAAGATGATGAGATGCTTAAGAAAGAACTCT 2272  
2168 -----GCMAAGCAAGAAAAAC---CAAGATGACATTTAAAGTTAGAAAGAACTCT 2218  
2273 GTGAAATATGTAAGAAAGATGGAAGAACGTGCAAAAGCTTACCAACAGAAATTAGAGAC 2332  
2219 GTGAAGGCTATTAAGCTATTAAGAAATAATGCTCAAAATTTGAAGACAGCTTGAAG 2278  
2333 AACCGCAGCTGATCTAAAAAAGATTAATGAATATGAGAACTTAAGAAACGTGAG 2392  
2279 AAGTAAAAAGAGCTCTTGAAGAAAGAAAGAAAGTATTAATAAATTTAAAGAAAGACAG 2338  
2393 AGGAAGCAATGAACAAGTCCAGTCTTGTGTTGTCACTCATTAAGAAAAACGAAGTATG 2452  
2339 AAAAAAGCATGAAGATGCAACCTTATTTTATGAGAGCGAAAGACCTGATTAATAT- 2397  
2453 TATCAAAAAGTAAATAGCAAAAACAAGATAAGATGCGGTTTCAAAACGACCTTCAAGATA 2512  
2398 -----AATATTAAGTCAATTAATTAAGACTCATGTGATACCTTAAGAGAAAGGCA 2452  
2513 CCACAAAACATGTGAATAATTAAGAGGAGGATTAAGAGTATTCGTTAACAGAAATTAAG 2572  
2453 CAACAGATTT-----AACTTGTAAAGAAATGCAAAAGTGCATTAACAGAAAAAG 2506  
2573 AAGCTAAAGCATTTGATTTGGCAGCAAGATTTTGAAGATATGTAGATTTGAAGAA 2632  
2507 AATTAGCAGCATTTGATTTGTACCAAGAGATTTGATCTTAATCTTAATTTGAAGAA 2566  
2633 GATGAATTAATTTGGAATCAATGATGCAAGATTAAGAGAGATTTGCAAAAGCTTAAGAAAG 2692  
2567 TATGTAATCATTTCACTGAAGATTTGTGTTTCAAAAAAGATG---TACTGTGAGGATC 2623  
2693 TATCAAAAAGATTAATTAAGCTTGTGCAATCTGAACCTTGAAGGTGAAGCGCAGC 2752  
2624 CATGTAATAAATACAGGGAATATGTTCAACATTAAGGCCACTAAAGTGAAGACACAGC 2683



```

Oy      1105  ACTTGCAAAATTTTCACCTTCAAAAGTCATAATATGCTCTTTTGATCCAAACAAAGGA 1164
Db      2199  AGATGTACAAAACCTCAAGTTAAATCTTATATGTGAAAGGGTGTGTCAGGAGTTAAAGAA 2258
Oy      1165  AATAATGAA-----ATTGTTGATGGGAGGGTGGCCACATTTCTTAGCAACGAAGAT 1218
Db      2259  GATAAAAAATCACATCTTTTATCGTGGGGAACGCTTCCACATTATTTATCGAAGGAGAG 2318
Oy      1219  TGTGCGAAATTTGAGTCTCTATTTGTTCTATTTTGAAAAAAAATGTCAGATGAGAAA- 1276
Db      2319  TGTGCGAGAACTTGAGTCCGAAATGTTTCTATTAGAAAAATGCGGTGAAGATATGAGAT 2378
Oy      1277  ----ATGATGTAAAAATATTAAGGACACATGTTACAAAAGAGACTTGATGACAGCGCA 1332
Db      2379  GGTGAAGGCTGTCAAAATCTACGATCACGCTGTATTAAGGACAAAGACAGATGTTG 2438
Oy      1333  AATAAAGTGTGCAAGAAATATATCGGAGGAATGTT---ACATGCTTCAACAAAGACTGG 1389
Db      2439  AATAAGTCTTTCAAAAGGAATGGAAGGAAAGCTTGCTATGTATGATTTATATAGCAT 2498
Oy      1390  CTTGAAAAGTTTCAACAAGAATTAGTAAAGTATGTGAAAATCTGAAAAAGAAAAACAA 1449
Db      2499  CCTAAAGATTGTAAAAAATATGTGTGTACAAAACGTACAAAACCTTAAAAAAGATTAAGA 2558
Oy      1450  GGAAGTTCTCAAAACGATGAATTATTATCTGTGTGTACAGCCAGCAAAAGACGCCGG 1509
Db      2559  TACCTTTCAAA-----TGTCTTTATCTTAAGAACTATGTTAT 2597
Oy      1510  TTGCTTACACATGATCTTGCATGAAAACTATCTTTTACGACAAACACTGGATCAAAAG 1569
Db      2598  GGGCTTTCAAAATGATATTTTCTCCAAATCCAAAGATTAAAGTTGCTTTTATGATGATCAG 2657
Oy      1570  CGAGATTTCCGACAGATAAAAATTTGCAGAAGATTGGGAGAGAAATGCCAAGATTAGGA 1628
Db      2658  AGAGATTTTTCATTGTAAGAAAGATTGTCTTGAAATTTGGAGAGAAATGTGATCAACTTGT 2717
Oy      1630  GAGGATTCAAAAGAAATATACATGGCCATGTCAATACATCAGTGAAGACAAATGCATG 1689
Db      2718  AGTATTTTCAATTATTCGATTTTACAAAAGTGTATATACATTTGAAAAGACGCTGTGATATTT 2777
Oy      1690  GGGACTACAGAAATTTTAAAGCAGGTTTAT 1721
Db      2778  GACGTTACAGAAAGATTAGAAAAGTATTTT 2809

RESULT 9
PCT-US92-08328-1
; Sequence 1, Application PC/TUS9208328
; GENERAL INFORMATION:
; APPLICANT: Fishman, Jay A.
; TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08328
; FILING DATE: 19920930
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/781,034
; FILING DATE: 18-OCT-1991
;

```

	Prior Application Data:	
	Application Number: US 07/766,166	
	Filing Date: 30-Sep-1991	
	Attorney/Agent Information:	
	Name: Granahan, Patricia	
	Registration Number: 32,227	
	Reference/Docket Number: MGH91-02AA PCT	
	Telecommunication Information:	
	Telephone: (617) 861-6240	
	Telefax: (617) 861-9540	
	Telex: 951794	
	Information for Seq ID No: 1:	
	Sequence Characteristics:	
	Length: 2814 base pairs	
	Type: Nucleic Acid	
	Strandedness: double	
	Topology: linear	
	Molecule Type: cDNA to mRNA	
	Hypothetical: NO	
	Anti-Sense: NO	
	Feature:	
	Name/Key: CDS	
	Location: 1617..2813	
	Pct-US92-08328-1	
	Query Match	8.5%; Score 261.6; DB 5; Length 2814;
	Best Local Similarity	51.3%; Pred. No. 1.3e-47;
	Matches	879; Conservative 0; Mismatches 764; Indels 69; Gaps 9;
Oy	25 GCACGAGGACACAGATATGACATTGATGAGAGAACATGTTTTACGCTTAATTCMAAGAA	84
Db	1152 GTACAAATGACACACAAGTAGATTTAAGAGAGAACACTTTTGCTTTCAATTGTGAAGAC	1211
Oy	85 GATGACTTAATGTAGCAGCATGCACAAAAAAAATAAAAAATAATGCCAAGAATTGACT	144
Db	1212 AAACATGATGATGATGATGATGATGCAAAAAAAGCGTCGAGGAATTTGTAAAGATGAG	1271
Oy	145 GAAGCAAACCTAAATATAGAACAAATCACAGAAAATCTTAAAGTTTTTGGCAAGATGCA	204
Db	1272 AAGACCATGATGCAATTT---CAGTGTGAATGAGAAAGTTAAAGACCTTGTGATGATGAAA	1328
Oy	205 AAACCATATCAAAATGCCAAGAACTGAAGCCCAATTGAGAAAAAATGACTACATC	264
Db	1329 AAAAGAACGAAAAAATCCAAAGAACTGAAAAAAGTAAAGTGAATGGCAACTTTT	1388
Oy	265 AAAGAAATCTTAAAGAGCAATTTAAAAAATTGACATTTAATACGATTAAGATTGC	324
Db	1389 GATACGATCTTGAGACATCGT-----AGATGACATGAAAGATGAAGAGTTGT	1439
Oy	325 AAAGAGATGAACAACATSCCTATTTTTGGAGGAGTATGTTCAAAAGACTTAAAGAT	384
Db	1440 AAAAAACATGAAGAAAATGTATCTTTTAGAGGACCAACCATAATGCTTAAAGAG	1499
Oy	385 GATTGCAATCTTTGAAAAATTAAGTGTATTCMAAAGAAAGTGATTAAGTTTCCGGAAGA	444
Db	1500 AACGTGTCTCAAGTTGAGGGAAGATGTTACGAATTTGAAGGATTAAGAGTGGCAGAGAG	1559
Oy	445 GTTCTTTTAAAGACACTTCGATGAGCTTTAATGATCAGTCAATGTGAAAAAAACTT	504
Db	1560 CTCCTTTTGAAGGCGCTCGAGGGGATGCTTAAGATGAAGCTAATGTTAAAGAAAAGTG	1619
Oy	505 AAAGAGATTTGCCCTGTCAATGGGAGGAAAGTATGATTAACAACTTGTGTCTGAC	564
Db	1620 AAAAATGTTGGCCCAATGTTAAGCCGAGAAAGTGAACAGCTGATGTTTCTCGCTGAT	1679
Oy	565 CAGAAAGACATGTAAGATATTTTAAATGAAAAAAGATAAGAAAGTCCGATCTCTTAA	624
Db	1680 TCGATGGAACGTGTAAAGCGCTGAAAAAACAACAAAGAAAGTTTCCCTGCTTTAAA	1739
Oy	625 ACAGATGTTTCAGACGACCTAGGAAGTTTAAAAAAGAAACATGCTTGAATTAAGTGA	684
Db	1740 GAAAAAGCTTAAAGTGGCGCAATT-----AAAGGAAAAAATGTCATGAAGACTTGAG	1790



QY 685 CAATGCTATTTTACATTTGCAAAATTCGCGAGACGACATATTAATTAATGATTAATG 744  
 DB 1791 AAATGCTATTTTTCACAAAGAGCTGTACTCA-----ACAAAGTGTGATGAGAT 1841  
 QY 745 GAGAGAAATTCGCAAGAACAAATATGCTTATATGCAACGAGACCCGATTTTGATCCA 804  
 DB 1842 ATGAAAGCAATGCAAGAAAAGATTCACATATTAAGCGCGAATCTGATTTAGTCT 1901  
 QY 805 ACTAGGCCAGAGGCTACATAGCGAGAGATATAGGCTGCGAAGATTTATTAAGAAGTA 864  
 DB 1902 GTCAAGCGAAGGGGTGGTGTGTGAGAAAGTATGGGTGGATGTGTATTAAGAAAGCT 1961  
 QY 865 GAGAGAGATGAGATTTTATTTGAAAGATCATCTAAGATGCGACAGCTTTGTTGCA 924  
 DB 1962 GAAAAAGAGAAATTTATTTTGAAGAAATCAGAGATGATCTACCAAGAACTGAGTACA 2021  
 QY 925 TTGTTGATCCAGATTTCTAGTCTTAAAAAAGACGACAAAGAAATGCGAAGAACCC 984  
 DB 2022 AAATTTCTGCAAGATCTCTGCTACTGTGAGCAGAGATGAGATGATGCGAGGAGAAA 2081  
 QY 985 CTTCACAAAAGCTGCAAAAATCCCTCATGACATGAGCTTTAGAAAGTTATATGAAGAA 1044  
 DB 2082 TGGCGTAAAGGGTGAAGAAATGAAACCTTCTAAGATTTTGAATCTGATTTGATGAG 2141  
 QY 1045 AATGTTTAAATGATATGATGAAACGAAAAATGTGAAAGATTTGCAAAATGATATTAACAAA 1104  
 DB 2142 TTATGCAAAAGTGTGATTAAGAAATTAATGCAAAAAAAGCTAGATG---TAAAGAA 2198  
 QY 1105 ACTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTTTTGTGATCCAAACAAAAGA 1164  
 DB 2199 AGATGTCACAAATCTCAAGTAAATCTTATGTGAAAGGGTGTCTACGAGATTTAAAGAA 2258  
 QY 1165 AATTAATGA-----ATTGTGATGAGGAGGGTTCGCAACATTTCTTAGACAGAAAT 1218  
 DB 2259 GATTAATAATCACTCTTTTATCGTGGGACAGCTTCCAACTATTAATTAACAAAGGAGAG 2318  
 QY 1219 TGTCGAAATTTGAGTCTATTTGTTTCTATTTTGAAGAAAAATGTCAGATGAGAAA-- 1276  
 DB 2319 TGTCGAAATTTGAGTCTATTTGTTTCTATTTTGAAGAAAAATGTCAGATGAGAAA-- 1276  
 QY 1277 ---ATGATGTAATAATTAAGCAACATGTTACAAAGAGACCTTGATGACGAGCA 1332  
 DB 2379 GGTGAGCGTGTCAAAATCTACGATCAGCGTATTAATAAGGACAAAGACAGATGTTG 2438  
 QY 1333 AATTAAGTCTGCAAGAAATATGCGAGATGTT--ACATGTTCAACAAAGCTGG 1389  
 DB 2439 AATTAAGTCTTCAAAAGAAATTAAGGAAAGCTTGTCTATGATTAATTAAGGAT 2498  
 QY 1390 CTTGAAAGTTTCAACAAAGATTAAGTAAGATATGTAAGAACTGAAGAAAGAAACAAA 1449  
 DB 2499 CTTAAGATTTGTAATAATTAAGTGTGTAAGAACTGTACAAACCTTAAAGAAAGTAAGA 2558  
 QY 1450 GGAAGTTCTCAACGATGATTAATTTATCTGTGTGTACAGCCAGCAAAAGACCCGG 1509  
 DB 2559 TACCTTTCAAAA-----TGTCTTATCCTAAGAACTATGTTAT 2597  
 QY 1510 TTGCTTACACATGATTTTCAATGAAAATCTATTTTACACAACTGATCAAAAG 1569  
 DB 2598 GGGCTTTCAAAATGATTTTCTCCAAATCCAAAGAGTTAAGTTGCTTTTGAATGATCAG 2657  
 QY 1570 CGAGATTTCCGCAAGATTAATAATTTGCAAGAAATTTGGGGAAGAAAGTCCAGATTTTGA 1629  
 DB 2658 AGAATTTTCAATTTGAAAGAGATTTGTTGTAATTTGGAGAGAGATTTGATCAACTTATG 2717  
 QY 1630 GAGATTTCAAAAGAAATTAACATGCGCATGATACATGAGACAGCAATGCAATGCGTTG 1689  
 DB 2718 AGTATTTCAATTTGATTAAGAAAGTATTAACATTTGAAGAAAGCGGTGATTAATTT 2777  
 QY 1690 GGGATCAAGAAATTTTAAAGCAGTTTAT 1721  
 DB 2778 GACGTTACAGAAAGATTTAAGAAAGTATTTT 2809

RESULT 10  
 US-07-781-034-3  
 : Sequence 3, Application US/07781034  
 : Patent No. 5442050  
 : GENERAL INFORMATION:  
 : APPLICANT: Fishman, Jay A.  
 : TITLE OF INVENTION: Molecular Cloning of Antigens Shared By  
 : NUMBER OF SEQUENCES: 4  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 : STREET: Two Militia Drive  
 : CITY: Lexington  
 : STATE: Massachusetts  
 : COUNTRY: USA  
 : ZIP: 02173  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/07/781,034  
 : FILING DATE: 19911018  
 : CLASSIFICATION: 424  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Granahan, Patricia  
 : REGISTRATION NUMBER: 32,227  
 : REFERENCE/DOCKET NUMBER: MGH91-02A  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (617) 861-6240  
 : TELEFAX: (617) 861-9540  
 : INFORMATION FOR SEQ ID NO: 3:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1448 base pairs  
 : TYPE: NUCLEIC ACID  
 : STRANDEDNESS: double  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: DNA (genomic)  
 : HYPOTHEICAL: NO  
 : ANTI-SENSE: NO  
 : US-07-781-034-3  
 Query Match 6.2%; Score 192; DB 1; Length 1448;  
 Best Local Similarity 54.1%; Pred. No. 1.9e-32;  
 Matches 565; Conservative 0; Mismatches 440; Indels 39; Gaps 7;  
 QY 35 CACAGATACCATTTGATGAGAAACATGTTTATGCTTAAATTTCAAGAAAGATGACTAA 94  
 DB 432 CACAGATGATTTAAGAGAGAACACTTTTGGCTTTCATTGTAAGAGACAAACATGATG 491  
 QY 95 GTGACGAGAAATGCAAAAAAACTAAAAAATTTGCCAAGATTAAGTGAAGCAAAAC 154  
 DB 492 ATGAGATGATGCAAAAAAGCTCGAGAAATTTGTAAGAGTTGAAGAAAGACATG 551  
 QY 155 TAAATATGAACAAGTACAGAAAACTTAAAGTTTTCGAAAGATGAGAAAGACAGTA 214  
 DB 552 AGAATTT---CAGTGAATGAGAAAGTTAAAGACTTTTGATGATTAATAAAGAGAGCG 608  
 QY 215 CAAAATGCAAGAACTGAAGCCCAATATTTGAGAAAAATTTACTACATCAATCAAGAAAC 274  
 DB 609 AAAAATGCAAGAACTGAAGAAAAAAGTTGGGATGAAATTTGGAACTTTGATACGATC 668  
 QY 275 TTAAGAAGCAATTAATAAAAAAATTTCAATATTAACGATTAAGATTTGCAAGAGAAATG 334  
 DB 669 TTGAAGCATCGGT-----AGATGACATAGAAAGTGAAGAGTTTGTAAAGAAACATG 719  
 QY 335 AACCAATGCGCTATTTTGGAGGAGTATGTTCAAAAGAACTTAAGATGATTCATTA 394  
 DB 720 AAGAAAAATGATTAATTTTGAAGAGAGAGCCCAATTAAGTTTAAGAGAACTGTGTCA 779  
 QY 395 CTTGAGAAATTAAGTCTATCAAAAGAAAGTGAATTAAGTTGCGAAGAAAGTTCTTTAA 454

```
Db 780 AGTTGAGGAAAGATGTTACGAATTGAAGCCGTAAGGCGGAGAGAGCTCTTTTGA 839
Oy 455 GAGCACTTCGATGCGATCTTAAATGATCATATGTGAAAAAACTTAAAGAGATT 514
Db 840 GGGCGCTCGGAAAGAGACTTAAAGAAAGTTAAATGTAAGCAGATGAAAAAGTTT 899
Oy 515 GCGCTGTCAATGGGAGGAAAGTGTAGTTAACAACTTGTGTGTAACGAGAAAGGA 574
Db 900 GCCCAGTGTAAAGCCGAGAAAGCGAGATGATGTTTGTGCTTGAATTCGATGGA 959
Oy 575 CATGTAAAGATATTTAAATGAAAAAGATTAAGAGTGGTACTTTAAACAGATGTT 634
Db 960 CGTGTGAG---CGTGAAAAAAAATTCAGAGAAATTTGCCAGCTTTTAAAGAAAGCTTA 1016
Oy 635 CAGCAGCACTAGGAAGTTTAAAGAAACATGTCTTGAATTAATCTGACACATGCTATT 694
Db 1017 AAGATG-----GCGAATTAAAGAAAAATGTCAATGAAGACTTGAGAAATGTCAAT 1067
Oy 695 TTTACATTGGAATTTGCGGAGACGATTAATTAATTAATTAATTAATTAATTAATTAAT 754
Db 1068 TTTAC-----GAGAAAGCGTGTGATTAACAAAAATGTGATGAGATTAAGATCAAT 1118
Oy 755 GCCAAGAACAAAAATATGCTTATATGCCACAGACCCGATTTGATCCAACTAGCCAG 814
Db 1119 GCGAGAAAAAGAAATCACAATTAAG---GCGCAAAATCTGATTTCTCTGCAAGCCGA 1177
Oy 815 AGGCTACATAGCAGAGATATAGGCTTGAAAGCTTTTATTAAGAAAGTAGAGAGATG 874
Db 1178 AGAGCTCGTGTGTGAAGATATGGGTTGATGATGTGTATTAAGAGCTGAAAAAGAA 1237
Oy 875 GAGTTTATTTGGAAGATCATCTAAGAGATGCCACAGCTTTGTGGCATTTGTGATCC 934
Db 1238 GAAATTAATTTAGAAATCAGAGATGATCTAACAGAGAGTCAAAATTTCTGC 1297
Oy 935 AAGATT-----CTAGCTTTAAAAAAAAGACGAAAGAAATGCGAAGAAAGCCCTCA 989
Db 1298 AAGATCTCTGTAGTCTTTGAGCAGATGAGATGATGATGATGATGATGATGATGATG 1357
Oy 990 AAAAAAGTGCAAAAATCTCATGAACATGAGGCTTTAGAAAGTTATGTAAAGAAATGG 1049
Db 1358 AAAAAAGCTTAAAAAATGTGAAACTTCAATGATTTGAATTAATGATTTGATGATGAT 1417
Oy 1050 TTTAAGTATGATGAGACGAAAA 1073
Db 1418 CAAGATGCTGATTAACAAAAAA 1441

RESULT 11
PCT-US92-08328-3
; Sequence 3, Application PC/TUS9208328
; GENERAL INFORMATION:
; APPLICANT: Fishman, Jay A.
; TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
; TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08328
; FILING DATE: 19920930
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/781,034
```

```
; FILING DATE: 18-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,166
; FILING DATE: 30-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MGH91-02AA PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; TELEX: 951794
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US92-08328-3

Query Match 6.2%; Score 192; DB 5; Length 1448;
Best Local Similarity 54.1%; Pred. No. 1.9e-33; Indels 39; Gaps 7;
Matches 565; Conservative 0; Mismatches 440;

Oy 35 CACGAATAGCATGATGAGAACATGTTTACCTTAATTTCTAAGAAAGATGACTTA 94
Db 432 CACAAGATGATTAAGAGAGAACACCTTTGCTTCATTTGATGAAGACAAACATGATG 491
Oy 95 GTGACGAGAAATGCAAAAAAATTAATAATTTGCCAAGATTTGACAGCAAAAC 154
Db 492 ATGAGAAATGATGCAAAAAAGCTCGAGAAATTTGTAAAGAGTTGAAAGAACAGATG 551
Oy 155 TAAATATGAACAAGTACACAGAAACCTTAAGGTTTGTGCGAAGTAGAAAGACAGATA 214
Db 552 AGAATTT---CAGTGTGAATGAGAAAGTTAAAGACCTTTGTGATGATTAATAAGAGACG 608
Oy 215 CAAATATCAAGAACTGAAGCCAAATATGAGAAAAATGTACTACATCAATCAAGAAAAAC 274
Db 609 AAAAAATCAAGAACTGAAAAAAAGTTGGGATGATTTGGAACTTTGTATCGGATC 668
Oy 275 TTAAGAAAGCAATTAATAAAAAAATTCAGATTATTAAGGATTAAGATTTGCAAAAGAAATG 334
Db 669 TTGAAGATCGGT-----AGATGACATGAAGATGAAGAGATTTGTAAAAAACATG 719
Oy 335 AACAAATATGCTATTTTGGAGGAGATGTCATAAAGAACTTAAGATGATTCATA 394
Db 720 AAGAAAAATGTATCTTTTGAAGAGACAGCCAAATAGTCTTAAGAGAACTGTCTCA 779
Oy 395 CTTTGAAGAAATAGCTATCAAAAAAGAAACGTGATTAAGTTGCGAAGAAAGTTCTTTTAA 454
Db 780 AGTTGAGGAAAGATGTTAGCAATTGAAGGTAATAAAGTTGGCAGAGAGCTCTTTTGA 839
Oy 455 GAGCACTTCGATGCGATCTTAAATGATCATATGTGAAAAAACTTAAAGAGATT 514
Db 840 GGGCGCTCGGAAAGAGCTTAAAGAAAGTTAATGTAAAGCAGAGATGAAAAAGTTT 899
Oy 515 GCGCTGTCAATGGGAGGAAAGTGTAGTTAACAACTTGTGTGTAACGAGAAAGGA 574
Db 900 GCCCAGTGTAAAGCCGAGAAAGCGAGATGATGTTTGTGCTTGAATTCGATGGA 959
Oy 575 CATGTAAAGATATTTAAATGAAAAAGATTAAGAGTGGTACTTTAAACAGATGTT 634
Db 960 CGTGTGAG---CGTGAAAAAAAATTCAGAGAAATTTGCCAGCTTTTAAAGAAAGCTTA 1016
Oy 635 CAGCAGCACTAGGAAGTTTAAAGAAACATGTCTTGAATTAATCTGACACATGCTATT 694
Db 1017 AAGATG-----GCGAATTAAAGAAAAATGTCAATGAAGACTTGAGAAATGTCAAT 1067
Oy 695 TTTACATTGGAATTTGCGGAGACGATTAATTAATTAATTAATTAATTAATTAATTAAT 754
```

Db 1068 TTAC-----CGAAGACGTGTGATTAACAAATGTGATGAGATTAAGATCAAT 1118  
QY 755 GCCAAGAAACAAATATGTGTATATGCAACGAGACCCGATTTTGATCCAATGAGCCAG 814  
Db 1119 GCGAGAAAGAAATACATATAAG--CGCCAAATCTGATTCGTCTCAAGCCGA 1177  
QY 815 AGCTACAAATAGCAGAGATATAGGCGCTGAAAGTTTATTAAGAGGTAGAGAGATG 874  
Db 1178 AGACGTGTTGTTGAGAAGTATGGGTTGGATGATGTGTATTAAGAGCTGAAGAAAG 1237  
QY 875 GAGTTTATTTGAAAGATCATCTTAAGAGTGCACGCTTTGGCATTTGTATCC 934  
Db 1238 GAATATATTTGAAATACAGAGTGTATCAACAGAGAGTCAAGTAATAATTTCTGC 1297  
QY 935 AAGATTT-----CTAGCTTAAACAAAGACAAAGAAAGAAATGCGAAGGCCCTTCA 989  
Db 1298 AAGATCTCTTGCTGATCTTGAGCAGAGTGAATGATTAAGAGTCAAGGAAAGAAATGCG 1357  
QY 990 AAAAGCTGCAAAATCTCTCAATGAGCTTTAGAAAGTTTATGTAAAGAAATGCG 1049  
Db 1358 AAAAGCTGTAAGAAATATGTAAGCTTCTAGTATTTGAATCTGATTTGATGAGTTATG 1417  
QY 1050 TTTAAGTATGATGAAACGAAAA 1073  
Db 1418 CAAAGATGCTGATTAACAAAAAA 1441

RESULT 12  
US-09-762-724-15  
; Sequence 15, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; PCT/US99/18750  
; PRIOR FILING DATE: 1998-08-17  
; PRIOR APPLICATION NUMBER: 60/096,805  
; PRIOR FILING DATE: 1998-08-17  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 249  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(249)  
US-09-762-724-15

Query Match 6.0%; Score 185; DB 4; Length 249;  
Best Local Similarity 83.9%; Pred. No. 3,6e-31;  
Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 2821 GAATGCAATCTTACAGACACACATGCGTTACACAGACATGACACACACAGC 2880  
Db 1 GAGTGCATCTCTGACAGACACACGCGGTCACAAAGACGTCGCCCATATACAGC 60  
QY 2881 ACCGCTCTATCATCTACATCAATCAAAATTAATGATGATGATGATGATGATGATG 2940  
Db 61 ACTTCTAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
QY 2941 AAACCAACCAAGTATGACAGACAGAGAGAGATGATGACAGAGACGTAACCGAGTAG 3000  
Db 121 AAGCTTGAAGATTTACAGACAGAGAGAGATGATGACAGAGAGAGAGAGAGAGAGAG 180  
QY 3001 GGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3060  
Db 181 GGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240

QY 3061 TTCATGATT 3069  
Db 241 TTCATGATT 249

RESULT 13  
US-09-902-540-1357  
; Sequence 1357, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1357  
; LENGTH: 612  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(612)  
; OTHER INFORMATION: unsure at all n locations  
US-09-902-540-1357

Query Match 2.9%; Score 87.6; DB 4; Length 612;  
Best Local Similarity 46.7%; Pred. No. 1e-09;  
Matches 273; Conservative 0; Mismatches 311; Indels 0; Gaps 0;

QY 83 AAGTGCATCTTACAGACACACATGCGTTACACAGACATGACACACACAGC 142  
Db 4 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 63  
QY 143 CTGAAGCAAACTTAATATGAAACAGTACACAGAAACTTAAGGTTTTCGCAAGATG 202  
Db 64 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 123  
QY 203 GAAAGCAAGATTAAGTAAAGAAAGTGAAGCAATTTAGAGAAAGAAAGTGTACTCAA 262  
Db 124 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 183  
QY 263 TCAAGGAAAGCTTAAGAAAGCAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 322  
Db 184 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 243  
QY 323 GCAAGCAAGTGAACAAATGCTTATTTTGAAGGAGATGTTCAGAAAGAAAGTAAAG 382  
Db 244 AAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 303  
QY 383 ATGATTCATATCTTGAAGAAATAGTGTATCAAGAAAGAAAGTAAAGTTCGGAAG 442  
Db 304 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 363  
QY 443 AAGTCTTTTAAAGCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 502  
Db 364 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 423  
QY 503 TTAAGAGATTTGCTCTGATGAGGAGGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 562  
Db 424 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 483  
QY 563 ACCGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 622  
Db 484 AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 543  
QY 623 AAACAGATGTTTCAGCAGCACTGAAAGTTTAAAGAAAGAAAGCA 666

Db 544 AA 587

RESULT 14  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpc-F18  
; US-08-232-463-14

Query Match 2.8%; Score 86.6; DB 1; Length 7218;  
Best Local Similarity 2.8%; Pred. No. 3.8e-09;  
Matches 11; Conservative 254; Mismatches 128; Indels 0; Gaps 0;

Db 56 AACATGTTTAACTTAAATCTAAAGAGATGACTAAGTGAAGAGAAATGCAAAAAA 115  
1456 AAGAGATGAAGAAATTTGGTACRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397  
116 AACTAAAAAATATGCGCAAGATGACTGAAGCAAACTAAATATAGACAAAGTACACA 175  
1396 RRR 1337  
176 GAAATCTTAAAGTTTGGCGAAGATGAAAGCATTAACAATGCAAGACTGAAG 235  
1336 RRR 1277  
236 CCAATATTTGAGAAAAATGACTACATCAAGAAAACTTAAAGACCAATTAATAAAAA 295  
1276 RRR 1217  
296 AAATTCAGATTATAACGATTAAGATTGCAAGAGAAAGAACAAATGCCATTATTTTGG 355

Db 1216 RRR 1157

Query 356 AGGAGTATGTTCAAGAAGCTTAAGATGATTCATACTTGAAGAAATAGTCTATC 415  
1156 RRR 1097  
416 AAAAGAAACGTGATTAAGTTGCGGAGAGAGTTTC 448  
1096 RRR 1064

RESULT 15  
US-09-902-540-1280  
; Sequence 1280, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(115849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1280  
; LENGTH: 1039  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(1039)  
; OTHER INFORMATION: unsure at all n locations  
US-09-902-540-1280

Query Match 2.4%; Score 74.4; DB 4; Length 1039;  
Best Local Similarity 46.2%; Pred. No. 9.4e-07;  
Matches 283; Conservative 0; Mismatches 323; Indels 6; Gaps 1;

Db 70 TTAATTCATAAGAGATGACTAAGTGAAGAGAAATGCAAAAAAATCTAAAAAATAT 129  
118 TTAATTAATAAAAAAATATAAAACAAATATTAATAAAAAAATATTAATAAAAAA 177  
130 TGCCAGATTTGACTGAAGCAAACTAAATATAGAACATACACAGAAACTTAAGT 189  
178 AATTAATAAAAAAATATAAAAAAATATAAAAAAATATAAAAAAATATAAAAAA 237  
190 TTTTGGAGATGGAAGACAGATACAAATGCAAGAGACTGAAGCCAAATATTGAGAA 249  
238 AATATATTAATAAAAAAATATAAAAAAATATAAAAAAATATAAAAAAATATA 297  
250 AAAATGTAATAATCAAGAGAACTTAAGAACCAATTAATAAAAAAATCAAGTTATA 309  
298 AAAAAAAAAAATATAAAAAAATATAAAAAATCAATCAATTAATAAAAAAATTAAT 357  
310 ACGATTAAGATTCATAAGAGATGAAACAATGCTATTTTGGAGAGATATGTTCA 369  
358 AATTAACAAAAAATATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATA 417  
370 AAAGAACTTAAGATGATTCGAATCTTTGGAATTAAGTCTATCAAAAGAAACGTAT 429  
418 AATTAATAAAAAAATATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATA 477  
430 AAAGTGGGAGAGAGTTCTTTTAAGACACTTGCTAGCGATCTTAAGATCACTATA 489  
478 AATTAATAAAAAAATATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATA 537  
490 -----TGTAATAAAAAAATTAAGAGATTGGCTGTGATGGAGGAGAAAGTATGAG 543  
538 AAAAAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATAATACATAT 597

```
QY      544  TTACAACTTGTCGTGAACCGAAGAGACATGTAAATATTTTAATGAAAAAGAT 603
      598  TTTAAAAACAAAAAACACCCGCTAAATATTAATACTATAAAAAAC 657
QY      604  AAGAAAGCGCGTACTCTTAAAAACAGATGTTTCAGCAGCACTAGAGAGTTTAAAAAGAA 663
      658  AAAAAAAGTGTCTTTCTTAAAAAAATCTCCCTTAAAAACATAAAAAAAA 717
Db
QY      664  ACATGCTTGAA 675
      718  AAAAGAAATTAA 729
Db
```

Search completed: October 29, 2005, 09:09:44  
Job time : 718 secs

*This Page Blank (uspto)*







Db 2101 CTCMAAGTAGAGCTTCAAGAAAAATTGACGTATTAATCTMAATGTGAACCTGATCGA 2160  
Qy 2161 AGATATTGTACAGTAGGGGAGAAAGTAAATAATGTGCTCAATCAGGGCTTAATGCAAGCT 2220  
Db 2161 AGATATTGTACAGTAGGGGAGAAAGTAAATAATGTGCTCAATCAGGGCTTAATGCAAGCT 2220  
Qy 2221 AACACCAAGATTAATCTGTGAAGAGTATGAGATGCTAGAAAAGAACTCTGTGAGAAA 2280  
Db 2221 AACACCAAGATTAATCTGTGAAGAGTATGAGATGCTAGAAAAGAACTCTGTGAGAAA 2280  
Qy 2281 TTAAGTGAAGAGTGAAGAAACAGTGCAGAACATTTACCAACAGATTGACAAACCGCA 2340  
Db 2281 TTAAGTGAAGAGTGAAGAAACAGTGCAGAACATTTACCAACAGATTGACAAACCGCA 2340  
Qy 2341 GCTGATCTAAAAAAGAAATTATAGACATATAGAGAACTTAAGAAACGGCAGAGAAAGCA 2400  
Db 2341 GCTGATCTAAAAAAGAAATTATAGACATATAGAGAACTTAAGAAACGGCAGAGAAAGCA 2400  
Qy 2401 ATGAAACAAGTCAGTCTTTTGTCTCACTCATTAAGAAAAAGAAAGTATGTATCAAAA 2460  
Db 2401 ATGAAACAAGTCAGTCTTTTGTCTCACTCATTAAGAAAAAGAAAGTATGTATCAAAA 2460  
Qy 2461 AGTAATAGCAAAAAACAGGATTAAGATCCGTTTCAACGAGCTTCAAGATCAACAAA 2520  
Db 2461 AGTAATAGCAAAAAACAGGATTAAGATCCGTTTCAACGAGCTTCAAGATCAACAAA 2520  
Qy 2521 CATGTGAAAAATATCTACGAGAGGGAGTTAAGATGTAATCCGTTAAGAAATTTGAAGCTTAA 2580  
Db 2521 CATGTGAAAAATATCTACGAGAGGGAGTTAAGATGTAATCCGTTAAGAAATTTGAAGCTTAA 2580  
Qy 2581 GCATTGTATTTGGCAGCAGAAATTTTGAAGATATGTAGATTTGAAGAAAGATGTAT 2640  
Db 2581 GCATTGTATTTGGCAGCAGAAATTTTGAAGATATGTAGATTTGAAGAAAGATGTAT 2640  
Qy 2641 AAATTTGAATCAGATTGCAAAATTAAGAGAGATTGCAAAAGCTTAAGAAAGATGTACAA 2700  
Db 2641 AAATTTGAATCAGATTGCAAAATTAAGAGAGATTGCAAAAGCTTAAGAAAGATGTACAA 2700  
Qy 2701 AAGTTAATAAGGCTTGTGCGAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACAGTG 2760  
Db 2701 AAGTTAATAAGGCTTGTGCGAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACAGTG 2760  
Qy 2761 ACAGAAAGTACAAAGCACTACAAACAAACAAACCCGTTGCCGATCCGAAGGCAACG 2820  
Db 2761 ACAGAAAGTACAAAGCACTACAAACAAACAAACCCGTTGCCGATCCGAAGGCAACG 2820  
Qy 2821 GAATGCAAAATCTTACAGACAAACAGACATGGGTTACACAGACATGCAACACAAAGC 2880  
Db 2821 GAATGCAAAATCTTACAGACAAACAGACATGGGTTACACAGACATGCAACACAAAGC 2880  
Qy 2881 ACGTCTACATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2940  
Db 2881 ACGTCTACATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2940  
Qy 2941 AAACCAACCAAGTGAAGCAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAG 3000  
Db 2941 AAACCAACCAAGTGAAGCAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAG 3000  
Qy 3001 GGGCTGAGAGATGAGTGGGTGAATGTGATGAGGGGGGATAGTGAAGCAATGTTATTCG 3060  
Db 3001 GGGCTGAGAGATGAGTGGGTGAATGTGATGAGGGGGGATAGTGAAGCAATGTTATTCG 3060  
Qy 3061 TTCAATGATTTAG 3072  
Db 3061 TTCAATGATTTAG 3072

RESULT 2  
US-10-654-416-11  
; Sequence 11, Application US/10654416  
; Publication No. US20040043409A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.

; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/10/654,416  
; PRIOR FILING DATE: 2003-09-02/762,724  
; PRIOR APPLICATION NUMBER: US/09/762,724  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3054  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3054)  
US-10-654-416-11  
Query Match 75.5%; Score 2318.8; DB 19; Length 3054;  
Best Local Similarly 86.0%; Pred. No. 0;  
Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;  
Qy 4 GCGCGGGGGGTCAAGCGGCAAGG---CAGCAGGCAACAGATATGATGAGAAACAT 60  
Db 1 GCGCGGGGGGTCAAGCGGCAAGG---CAGCAGGCAACAGATATGATGAGAAACAT 60  
Qy 61 GTTTTACCTTTAAATTTCAAGAAATGCACTAAGTGAAGAGAGATGAGAGAGATGAGAGAGAT 120  
Db 61 ATTTTGGCGTTGATTTCAAGAAAGATGCAATGAGAGATGAGAGAGATGAGAGAGAT 120  
Qy 121 AAAAAATATTGCCAAGAAATGCACTGAAGCAAACTTAATATGAAACAAAGTACAGAAAA 180  
Db 121 GAAAAAATCTCGAAGAGGTTGAAAAAAGCAATCACTAGAAATGAAAAAGTACATTAATG 180  
Qy 181 CTTAAGGTTTTCGCAAGATGCAAAAGCAGATCAAAATGCAAAAGTCAAAAGCCAAAT 240  
Db 181 CTTAAGGTTTTCGCAAGATGCAAAAGCAGATCAAAATGCAAAAGTCAAAAGCCAAAT 240  
Qy 241 ATTGAAAAAATGTACTACATCAATCAAAAGAACTTAAGAAAGCAATTAAGAAAAAATT 300  
Db 241 CAAGCAAAAGTTACGGGAAATGTACAAATTTTAAACAAAGCTAAGACCGCTTA 300  
Qy 301 CAGATTAAACGATTAAGATGCAAGAGATGCAAGAGATGCAAGAGATGCAAGAGATGCAAGAG 360  
Db 301 ACAATATCATCAGATGATTAATGCAAGAGAGATGCAAGAGATGCAAGAGATGCAAGAGAG 360  
Qy 361 GTATGTTCAAAAGAACTTAAGAGATGCAATCTTGAAGAAATGAAGTGAATGAAG 420  
Db 361 GCATG---CCATTAATCTTGAGAGATGTAACAACTTAAGAAATCTATGTATACAGAAA 417  
Qy 421 AAACGTATAAGTTGGGAAAGATCTTTTAAGACACTTCTAGACATCTTAATGGA 480  
Db 421 AAACGTATAAGTTGGGAAAGATCTTTTAAGACACTTCTAGACATCTTAATGGA 480  
Qy 478 ACAGAAACACATGAAAAAAGCTGAAGAGATTTGCCAGCTTTCAGAGGAGAAAGTAAAT 537  
Db 478 ACAGAAACACATGAAAAAAGCTGAAGAGATTTGCCAGCTTTCAGAGGAGAAAGTAAAT 537  
Qy 541 GAATTAACAACTGTGTGCAACCAAGAGAGATGTAAGATTTTAATGGAAGAA 600  
Db 541 GAATTAACAACTGTGTGCAACCAAGAGAGATGTAAGATTTTAATGGAAGAA 600  
Qy 598 GATTAAGAGTGGTACTCTTTAAACAGATGTTTCAAGAGATGAGAGATGAGAGATGAGAGAT 660  
Db 598 GATTAAGAGTGGTACTCTTTAAACAGATGTTTCAAGAGATGAGAGATGAGAGATGAGAGAT 660  
Qy 661 GAAACATGCTGTAATTAATGCAAGATGCTATTTTCAATGGAATTTGGAGAGAGAC 720  
Db 661 GAAATATGCTGTAATTAATGCAAGATGCTATTTTCAATGGAATTTGGAGAGAGAC 720  
Qy 717 GAAATATGCTGTAATTAATGCAAGATGCTATTTTCAATGGAATTTGGAGAGAGAC 717  
Db 717 GAAATATGCTGTAATTAATGCAAGATGCTATTTTCAATGGAATTTGGAGAGAGAC 717

OY 721 GATATTAATTAATGTAATGGAATGAGGAAATGCCAAGAACAAATATGCTATATG 780  
DB 718 GATATTAATTAATGTAATGGAATGAGGAAATGCCAAGAACAAATATGCTATATTA 777  
OY 761 CCACAGAGACCCGATTTTGAATCCACTAGCGAGGCGCTACAAATAGCAGAGATPAGG 840  
DB 778 CCACAGAGACCCGATTTTGAATCCACTAGCGAGGCGCTACAAATAGCAGAGATPAGC 837  
OY 841 CTGGAAGAGTTTATTAAGAGAGTAGAGAGATGAGATTTTATTTGAAGAGATCATCTA 900  
DB 838 CTGATGAGCTTTATTAAGAGAGTAGAGAGATGAGATTTTATTTGAAGAGATCATCTTA 897  
OY 901 AGAGATGCGACAGCTTTGTTGCGATTTGTTG--ATCCAGATTTAGTCTTAAAAAAA 957  
DB 898 AGAGATGCGACAGCTTTTATGCGTTGTTGTTAAGAAAGATGATACAGAAAAATATAT 957  
OY 958 GACGACAAAGAGAAATGCGAAGAGACCCCTCAAAAAAGCTCAAAAAATCCATGATAAT 1017  
DB 958 AATATCGGAGAAAAATGCAATTAAGATTTCCGAAAGATAAATGCAAAAACTTCAACAGAT 1017  
OY 1018 GAGCTTTAGAAAGTTTATGTAAGAAAAATGCTTTAAGTAATGATGAAACGAAAAATGT 1077  
DB 1018 GAGCTCTTAAGAAAAATTTATGTAATTAATATGTCCTATGCAATGAGAAAAAATGTC 1077  
OY 1078 GAAAGATTCGAAATGATATTAACAAAATTGCAAAATTTTCACTTCAAAAGTCACTAT 1137  
DB 1078 AAGAGATTTAGAAAGATATTTAAAAAAACATGCAAAACCTCAAAACGATCTTTAAA 1137  
OY 1138 AATGCTTTTGTGATCCAAACAAAGAAATATATGAAATTTGTTGAGGAAAGGTTGCA 1197  
DB 1138 AATGCTTTTATGATCC-----AATGATTAATTTGTTGAGGAGAAAACTGCA 1188  
OY 1198 ACATTTCTTAGACGAGAGATGTCGAAATTTGAGCTCTATGTTTCTATTTGAAAAA 1257  
DB 1189 ACATTTCTTACTTAATGAAGACTGTGCAAGATTTGATCTTATGTTTTTACTACAAAAA 1248  
OY 1258 AATGCTCAGATGAGAAAAATGCAATGTAAAAATATTAAGCAACATGTTACAAAAGAGA 1317  
DB 1249 GCTTGTCCAAATGCCAAAGAAAGATGTATGAAATTTGAGGCGCAGCTGTTATTAAGAGAGG 1308  
OY 1318 CTGTATGCGCGGGCAAAATTAAGTGTGCGCAAGAAATATGCGAGAAATGTTACA 1377  
DB 1309 CTGTATGCGCGGGCAAAATTAAGTGTGCGCAAGAAATATGCGAGTATTAATGCTGTCA 1368  
OY 1378 AACAAAGCTGCTTGAAAGTTTCAACAGAAATTAAGTAATGTAAGAGAACTGAAA 1437  
DB 1369 AATCAAGTTGCTTAAGAGATTTCAACAAAGATTAAGTAATGTAAGAGAGCT---A 1425  
OY 1438 AAAAGAAAACAAAGAAAGTTTCTCAACAGATGAATTTATTTCTGTGTGACAGCCAGA 1497  
DB 1426 AAAAGAAAATTAAGAAAGTTTCCCAACAGATGAATTTGTTCTGTGTGACAGCCAGA 1485  
OY 1498 AAAAGAGCCCGTGTCTTACACATGATCTTGAATGAAACATATCTTTTTCAGACAAG 1557  
DB 1486 AAAAGCTGACGATTTACTTACACAGATCTCAATATGAGGCTTATCTTTTACAGACAA 1545  
OY 1558 CTGTATCAAAAGCGAATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTGC 1617  
DB 1546 TTGATTCAAAAGAGAGATTTTCCGACAGATTAAGATCTGCAAGAGATTAAGGAAAAAATGTC 1605  
OY 1618 CAAGATTTTGAAGAGATTTCAAAAGAAATTTACATGCGCCATGCTATACACTGAGAGACAA 1677  
DB 1606 CAAGATTTTGAAGAGATTTCAAAAGAAATTTACATGCGCCATGCTATACCTGAGAGACAA 1665  
OY 1678 TGCAATTCCTTGGGAGCTACAGAAATTTTAAAGCAGTTTATTTGAGTGAACACAAAGAT 1737  
DB 1666 TGCAATTCCTTGGGAGCTACAGAAATTTTAAAGCAGTTTATTTGAGTGAACACAAAGAT 1725  
OY 1738 ACTTTGAAGAACAAAGAAAGTTGTGTAAATATCTTAAAGAAAGATGTAAATGCTCT 1797  
DB 1726 ACTTTGAAGAACAAAGAAAGTTGTGTAAATATCTTAAAGAAAGATGTAAATGCTCT 1785

OY 1798 AGAAGAGAGATGACCGTTTCTCTTTTGTATGTGTCTTCCAAAAACGCTACGTTGAGCTG 1857  
DB 1786 AGAAGAGAGATGACCGTTTCTCTTTTGTATGTGTCTTCCAAAAACGCTACGTTGAGCTG 1845  
OY 1858 ATGTTAAAAAGACGGAAGAGACAGGTGTGAAGTATTCAAAAAAAATATTAAGCTTCATAT 1917  
DB 1846 ATGTTAAAAAGACGGAAGAGACAGGTGTGAAGTATTCAAAAAAAATATTAAGCTTCATAT 1905  
OY 1918 ATTAATGAATTTCTTGAATAATTAATCAAAATTAATAACAACTGGAAGAAATTTGCTCC 1977  
DB 1906 ATTAATGAATTTCTTGAATAATTAATCAAAATTAATAACAACTGGAAGAAATTTGCTCC 1965  
OY 1978 TCTTGCGATACGTATTCGAAATGAATTTTCACTAATTTGCTCAAGCTTTACGAAAGAGAT 2037  
DB 1966 TCTTGCGATACGTATTCGAAATGAATTTTCACTAATTTGCTCAAGCTTTACGAAAGAGAT 2025  
OY 2038 AGTTGTCAAAAATCAAGAAAGCATCGTAGCGCTTCTATAAAGAAAGGCGCTTGGAAAGAT 2097  
DB 2026 AGTTGTCAAAAATCAAGAAAGCATCGTAGCGCTTCTATAAAGAAAGGCGCTTGGAAAGAT 2085  
OY 2098 GCTCTCAAAAGTAGAGCTTCAAGAAAAATTTGACTGATTAATCTAAATGTGAACCTGCAATTG 2157  
DB 2086 GCTCTCAAAAGTAGAGCTTCAAGAAAAATTTGACTGATTAATCTAAATGTGAACCTGCAATTG 2145  
OY 2158 AAAAGATTTGTACATGATAGCGGAGAAAGTAAATTAATGCGTAATCACTGCTTATGCAA 2217  
DB 2146 AAAAGATTTGTACATGATAGCGGAGAAAGTAAATTAATGCGTAATCACTGCTTATGCAA 2205  
OY 2218 GCTAACACCAAGATTAATCTCTGGAAGAGTAGATGCTGTAAGAAAGAACTCTGTAG 2277  
DB 2206 GCTAACACCAAGATTAATCTCTGGAAGAGTAGATGCTGTAAGAAAGAACTCTGTAG 2265  
OY 2278 AAATTTAGTGAAGAAAGTGAAGAAACAGTGCAAAGCATTTACCAAGAAATTAAGCAACCG 2337  
DB 2266 AAATTTAGTGAAGAAAGTGAAGAAACAGTGCAAAGCATTTACCAAGAAATTAAGCAACCG 2325  
OY 2338 GCACCTGATCTTAAGAAAAAGATTAATTAAGCATATAGAGAACTTAAAGAAACGTCAGAGAA 2397  
DB 2326 GCACCTGATCTTAAGAAAAAGATTAATTAAGCATATAGAGAACTTAAAGAAACGTCAGAGAA 2385  
OY 2398 GCAATGAACCAAGTCCAGCTCTGTTTGTCACTCACTTAAGAAAAAGAAAGTATGTATCA 2457  
DB 2386 GCAATGAACCAAGTCCAGCTCTGTTTGTCACTCACTTAAGAAAAAGAAAGTATGTATCA 2445  
OY 2458 AAAAGTAAATGCAAAAAACAAAGATTAAGAAATGCGCTTCAAAACGAACTTCAAGATACCA 2517  
DB 2446 AAAAGTAAATGCAAAAAACAAAGATTAAGAAATGCGCTTCAAAACGAACTTCAAGATACCA 2505  
OY 2518 AAACATGTGAATAATCTACGAGAGGAGTTAAAGATGTAATCCGTAACAGAAATTAAGAGCT 2577  
DB 2506 AAACATGTGAATAATCTACGAGAGGAGTTAAAGATGTAATCCGTAACAGAAATTAAGAGCT 2565  
OY 2578 AAACATTTGATTTGGCAGCAGAGATATTTGGAAGATGTATGATATTTGAAGAAAGATGT 2637  
DB 2566 AAACATTTGATTTGGCAGCAGAGATATTTGGAAGATGTATGATATTTGAAGAAAGATGT 2625  
OY 2638 AATAAATTTGAATCAGATTTGAGAAATTAAGAGAGATTTGCAAAAGCTTAAAGAAAGTATGC 2697  
DB 2626 AATAAATTTGAATCAGATTTGAGAAATTAAGAGAGATTTGCAAAAGCTTAAAGAAAGTATGC 2685  
OY 2698 AAAAAAGATTAATTAAGGCTTTGTCGCAATCTGAAGCTCTGAGAGGTGAAGCCGCAAGAAACA 2757  
DB 2686 AAAAAAGATTAATTAAGGCTTTGTCGCAATCTGAAGCTCTGAGAGGTGAAGCCGCAAGAAACA 2745  
OY 2758 GTGACAGAAATGACAAAGCAACTTAACAAACAAACAAACGTTGCGAGTCCGAAGGCA 2817  
DB 2746 GTGACAGAAATGACAAAGCAACTTAACAAACAAACAAACGTTGCGAGTCCGAAGGCA 2805  
OY 2818 ACGGAATGCAAAATCTTTACAGACAAACAGACACATGGTTTACAGACATTCGACACACACA 2877  
DB 2806 ACGGAATGCAAAATCTTTACAGACAAACAGACACATGGTTTACAGACATTCGACACACACA 2865  
OY 2878 AGCAGCTTACTATCACTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2937

|||||  
Db AGACGCTCTACTACATCTACATCCATCAATCAAAAAATTAATTGATGATCAAGAGCGA 2925  
Qy 2338 TGCACCAACCAAGTGTACGACGAGAGAGAAATGATGACAGACGTAAGCCAGT 2997  
Db 2926 TGCACCAACCAAGTGTACGACGAG-----GGATGATGCAAGAGACGTGAACCCAGT 2979  
Qy 2998 GAGGCGCTGAGATGAGTGGGTGGAATGTGATGAGGGGGGCTGATGATGCAATGCTTATT 3057  
Db 2980 GAGGCGCTGAGATGAGTGGGTGGAATGTGATGAGGGGGGCTGATGATGCAATGCTTATT 3039  
Qy 3058 TCGTTCATGATTTAG 3072  
Db 3040 TCGTTCATGATTTAG 3054  
  
RESULT 3  
US-10-654-416-7  
Sequence 7, Application US/10654416  
Publication No. US20040043409A1  
GENERAL INFORMATION:  
APPLICANT: Kovacs, et al.  
TITLE OF INVENTION: Identification of a region of the major surface  
FILE REFERENCE: 4239-58054  
CURRENT APPLICATION NUMBER: US/10/654,416  
PRIOR FILING DATE: 2003-09-02  
PRIOR APPLICATION NUMBER: US/09/762,724  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: PCT/US99/18750  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: US 60/096,805  
PRIOR FILING DATE: 1998-08-17  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 3084  
TYPE: DNA  
ORGANISM: Pneumocystis carinii sp. f. hominis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3084)  
US-10-654-416-7  
  
Query Match 73.7%; Score 2262.6; DB 19; Length 3084;  
Best Local Similarity 84.7%; Pred. No. 0;  
Matches 2640; Conservative 0; Mismatches 399; Indels 78; Gaps 6;  
  
Qy 1 ATGCGCGGGCGGTCAAGCGGCAGCAGCAGGACACAGATAGCATTTGATGAGAAACAT 60  
Db 1 ATGCGCGGGCGGTCAAGCGGCAGCAGCAGGACACAGATAGCATTTGATGAGAAACAT 60  
Qy 61 GTTTTACCTTTAATCTTAAGAGAAATGACTTAAGTGCAGAGATGCAAAAAAACTA 120  
Db 61 GTTTTACCTTTAATCTTAAGAGAAATGACTTAAGTGCAGAGATGCAAAAAAACTA 120  
Qy 121 AAAAAAATTTGCCAGAAATGACTGAGCAAAATCTTAATATAGAAACAAGTACAGAAAA 180  
Db 121 AAAAAAATTTGCCAGAAATGACTGAGCAAAATGACTGAGCAAAATGACTGAGCAAA 180  
Qy 181 CTTAAGGTTTTCGAAAGATGAAAGCAGATCAAAAATGCAAAAGAACTGAAGCCAT 240  
Db 181 TTTAAAGATTTCTGTGATACGGGAAACGAAATGAAATGTCAGATCTTAAGAAACAA 240  
Qy 241 ATTGAGAAAAATGTACTACATCAAGAGAAATCTTAAGAGCAATTTAAAAAAATTT 300  
Db 241 GTCAATCAAAAATGTACTAATTTCAAGAGAAATCTTAAGAGCAATTTAAAAAAATTT 300  
Qy 301 CAGATTATACGATTAAGGATTTGCAAAAGAAATGAAACAATGCTTATTTTGAAGGA 360  
Db 301 TCAGAAATTAAGATGAGATTTGCAAAAGAAATGAAACAATGCTTATTTTGAAGGA 360  
Qy 361 GTATGTTCAAAAGAACTTAAGATGATTCAGAAATCTTTGAGAAATTAAGTCTCAAAAG 420

|||||  
Db GCATGTCACAGAACTTAAGATGACTGCATTAATTAAGAAATACCTGTATCAAAAA 420  
Qy 421 AAACGTGATAAAGTTGGGAGAAAGTTCTTTTAAGACACTTGTGTACGATCTTAATGA 480  
Db 421 GAACGGAACAATGTGGAGAAAGTTCTTTTAAGAGGCGTTGTGTGTATCTCAATGA 480  
Qy 481 TCAATCATGTGTGAAAAAATTAAGAGATTTGCTGTCAATGGGAGGAAAGTAT 540  
Db 481 ACAAGACATGTGAAAAAATTAAGAGATTTGCTGTCAATGGGAGGAAAGTAT 540  
Qy 541 GAGTTAACAACTTGTCTGAACAGAAAGACATGTAAAGAAATTTTAATGAAAA 600  
Db 541 GAATTAACGAGCTTGTCTTTATCAAAAAACAATGCTAAGTCTTTGAACAAAAAGA 600  
Qy 601 GATTAAGATGCGGTACTCTTTAAACAGATGTTTCAGACACATAGAAATTTTAA--- 657  
Db 601 AAAAGTAAATGTATCTTGTGAAGAAAGATGAAAGACATTAAGAAATGAATTTG 660  
Qy 658 AAAGAAATGCTGTGAATTAATGCAATGCTATTTTACATTTGAAATTTGGGAG-- 715  
Db 661 CGAGAAAAATGTCTACTATTACTTTGACCAATGTTACTTTCACAGAGGAACTGTGAAGA 720  
Qy 716 -----ACGACATTAATTAATGT 735  
Db 721 GACAAATCAAGTCATTAACCTAATTAATTAAGTCAAGAAATATGTACAGAGTGT 780  
Qy 736 ATTGAATTTGGAGAAAAATGCCAAGAAACAATTTGCTTATATGCCACAGGACCCGAT 795  
Db 781 GATGAATTTACAGAAAGTGTGAAAAAATTTGTTATATGTCATCCAGATCCGAT 840  
Qy 796 TTTGATCCAACTAGGCGGAGGCTACATAGCAGAGATTAAGGCTGGAAGATTTAT 855  
Db 841 TTGATCACTAATGACGAGCTTACCTAGCAGAGACATAGGCTGGAAGATTTAT 900  
Qy 856 AAGAGGTAGAGAGATGAGATTTTATTTGAAAGAAATCACTAAGATGCGACAGT 915  
Db 901 AAGAGGCAAGAGATGAGATTTTGTGAAAGACAAATGTAAAGATGCAACAGCT 960  
Qy 916 TTGTGGCATTGTGTATCCAAAGATTCAGTCTTAATAAAAAAAGCAGCAAAAGAAATGC 975  
Db 961 TTGTGGCATTGTGTATCCAAAGATTCAGTCTTAATAAAAAAAGCAGCAAAAGAAATGC 1005  
Qy 976 GAAGAGCCCTTCAAAAAGCTGCAAAAATCCATGATGAGAGGCTTTAAGAAAGTTA 1035  
Db 1006 ATTAAGCCCTTCAAAAAGCTGCAAAAATCCATGATGAGAGGCTTTAAGAAAGTTA 1065  
Qy 1036 TGTAAAGAAAAATGTTAATGATGAGAACGAAAAATGTGAAGATTTGCAAAATGAT 1095  
Db 1066 TGTAAAGAAAAATGATGATGAGAACGAAAAATGTGAAGATTTGCAAAATGAT 1125  
Qy 1096 ATTAACAAAATTTGCAATTTTCACTTCAAAAGTCACTAATTAATGCTTTTGTATCCA 1155  
Db 1126 GTTAACAAAATTTGCAATTTTCACTTCAAAAGTCACTAATTAATGCTTTTGTATCCA 1185  
Qy 1156 ACAAAAGAAATATGAAATTTGATGAGAGGCTGCAACATTTCTTAGCAAGAA 1215  
Db 1186 CC-----TGATGAAATTTGCGAAATGCGAAATTTACCGACATTTCTTAGTATGAA 1236  
Qy 1216 GATTGTGCAAAATTTGAGTCTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1275  
Db 1237 GATTGTGCAAAATTTGAGTCTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1296  
Qy 1276 AATGATGTTAAATATTAAGCAATTTTCAAAAGAGACTTATGTCACGGGCAAT 1335  
Db 1297 GAAGCTTGTATGATGAGAGGCGCTGTTTACAAAGAGAGGCTTGTATGACGGGCAAC 1356  
Qy 1336 AAGTGTGCAAGAAATATGCGAGGAATTTCAATGTTCAACAAAGCTGCTGAA 1395  
Db 1357 AGTGTGTGCAAAATATGCGGCTTATTTAGTGTCTCAATTTCAAGTTGCTTAAG 1416  
Qy 1396 AAGTTCAACAGAAATTAAGATGATGTGAGAAATCTGAAAAAGAAACAAAGAGT 1455

```

Db 1417 GAGTTTCAACAAAGATTAGTAAAGTATGTAGAGCT---AAAAGAAATTAAGAAAGT 1473
Oy 1456 TTCTCAAA CGATGAATTAATTTATCTGTGTGTACAGCCAGCAAAAAGACCCGGTTGCTT 1515
Db 1474 TTCCCAACGATGAATATTTGTTCTGTGTGTACAGCCAGCAAAAAGCTGACGATTACTT 1533
Oy 1516 ACACATGATCTTCGATGATGAAAACTATCTTTTACGACAAACAACTGATTCAAAAGGAAAT 1575
Db 1534 ACACACGATCATAAATGAGGTTACCTTTTACGACAAACAAATGGATCCAAAAGGAAAT 1593
Oy 1576 TTCCCGACAGATTAATAAATTTGCAAGGAATTTGGGAGAAAAGTCCCAAGATTTTAGAGAGAT 1635
Db 1594 TTTCGACAGATTAAGACTGACAGAGAACTAGGGAATAAATGCAAGATTTTAGAAGAGAT 1653
Oy 1636 TCAAAAAGAAATTAACATGCGCATGTATACATGAGCAGCAATGCAATGCGTTGGGAACT 1695
Db 1654 TCAAAAAGAAATTAACATGCGCATGTATACATGAGCAGCAATGCAATGCGTTGGGAACT 1713
Oy 1696 ACAGAAATTTTAAACAGGTTTATTTGATGACAAAGATACCTTTGAAAGACCAAGAA 1755
Db 1714 ACAGAAATTTTAAACAGGTTTATTTGATGACAAAGATACCTTTGAAAGACCAAGAA 1773
Oy 1756 AGTTGTGTAATATACCTAAAAGAAAGTGTAAATATGCTCTAGAAGAGAGATGACCGT 1815
Db 1774 AGTTGTGTAATATACCTAAAAGAAAGTGTAAATATGCTCTAGAAGAGAGATGACCGT 1833
Oy 1816 TTCTCTTTGTATGTGTCTTCCAAAACGCTACGTGTGAGCTGATGTTAAAGACGTGAA 1875
Db 1834 TTCTCTTTGTATGTGTCTTCCAAAACGCTACGTGTGAGCTGATGTTAAAGACGTGAA 1893
Oy 1876 GACAGGTGTAAGTATTCAAAATAAATATTAAGATCTATATATTAATGAAATTTCTTAA 1935
Db 1894 GACAGGTGTAAGTATTCAAAATAAATATTAAGATCTATATATTAATGAAATTTCTTAA 1953
Oy 1936 AATAATACAAATTAATAATAACAACCTGGAAGAAGATTTCCCTCTTGACATACGTAATTC 1995
Db 1954 AATAATACAAATTAATAATAACAACCTGGAAGAAGATTTCCCTCTTGACATACGTAATTC 2013
Oy 1996 AATAGATTTTCACTAATTTGTCAGAGCTTTACGAAAAGAAATGTTGTACAAAATCAAG 2055
Db 2014 AATAGATTTTCACTAATTTGTCAGAGCTTTACGAAAAGAAATGTTGTACAAAATCAAG 2073
Oy 2056 AAGCATCGTAGCCGCTCTATTAAGAAAGCCCTGGAAGATGCTCTCAAAAGTAAAGCTT 2115
Db 2074 AAGCATCGTAGCCGCTCTATTAAGAAAGCCCTGGAAGATGCTCTCAAAAGTAAAGCTT 2133
Oy 2116 CAAGGAAATTTGACGTATTAATCTAAATGTGAACTGTGTAAGAAAGATTTGTACAGTA 2175
Db 2134 CAAGGAAATTTGACGTATTAATCTAAATGTGAACTGTGTAAGAAAGATTTGTACAGTA 2193
Oy 2176 GCGGGAACGTAAATATATGCGTCAATCAGTGGCTTATGCAAAAGCTAACCAAGATTAAC 2235
Db 2194 GCGGGAACGTAAATATATGCGTCAATCAGTGGCTTATGCAAAAGCTAACCAAGATTAAC 2253
Oy 2236 TCTGAAAGAGTAGAGAGATGCTAGAAAGAACTCTGTGGAATTAATGTGAANAAGTGTG 2295
Db 2254 TCTGAAAGAGTAGAGAGATGCTAGAAAGAACTCTGTGGAATTAATGTGAANAAGTGTG 2313
Oy 2296 GAAAGACAGTGCAGAACTTAACCAACAGATTTAGCAACAACGCGACGTGATTAATAAAA 2355
Db 2314 GAAAGACAGTGCAGAACTTAACCAACAGATTTAGCAACAACGCGACGTGATTAATAAAA 2373
Oy 2356 GATTATTAAGACATATGAGAACTTAAGAAACGTGCAAGAGAAAGCATGAACAAAGTCCAGT 2415
Db 2374 GATTATTAAGACATATGAGAACTTAAGAAACGTGCAAGAGAAAGCATGAACAAAGTCCAGT 2433
Oy 2416 CTGTGTTTGTCACTGATTAAGAAAAACGAAAGTAAATGTAATTAAGTAAAGTAAAGTAA 2475
Db 2434 CTGTGTTTGTCACTGATTAAGAAAAACGAAAGTAAATGTAATTAAGTAAAGTAAAGTAA 2493
Oy 2476 AAGATTAAGATGCGGTTTCAACGAGACTTCAAGATTAACCAAAAACATGTGAAATATACCTA 2535
Db 2494 AAGATTAAGATGCGGTTTCAACGAGACTTCAAGATTAACCAAAAACATGTGAAATATACCTA 2553

```

```

Oy 2536 CGAGCGGAGTTAAGATGTATCCGTAAACAGATTTAGAGCTTAAGCAATTTGATTTGGCA 2595
Db 2554 CGAGAGAGAGTTAAGATGTATCCGTAAACAGATTTAGAGCTTAAGCAATTTGATTTGGCA 2613
Oy 2596 GCAGAGATTTTGAAGATATGTAGATTTTGAAGAAAAGATTAATTAATTTGAATTCAGAT 2655
Db 2614 GCAGAGATTTTGAAGATATGTAGATTTTGAAGAAAAGATTAATTAATTTGAATTCAGAT 2673
Oy 2656 TGCAAGATTTAAGAGATTGCAAAAGCTTAAGAAAGATATGCAAAAAGATTAATTAAGGCT 2715
Db 2674 TGCAAGATTTAAGAGATTGCAAAAGCTTAAGAAAGATATGCAAAAAGATTAATTAAGGCT 2733
Oy 2716 TGTGCAATCTGAAGCCTCTGAGAGTGAAGCCGCAACGAAACAGTGAAGAAATGACAACG 2775
Db 2734 TGTGCAATCTGAAGCCTCTGAGAGTGAAGCCGCAACGAAACAGTGAAGAAATGACAACG 2793
Oy 2776 ACAACTCAACAAACAAACAAACGCTTGGGATCCGATCCGAAGGCAAGGAATGCAAAATCCCTTA 2835
Db 2794 ACAACTCAACAAACAAACAAACGCTTGGGATCCGATCCGAAGGCAAGGAATGCAAAATCCCTTA 2853
Oy 2836 CAGACAAACAGACATGGGTTACACAGACATGACACACACAAACGCTTACTATACACA 2895
Db 2854 CAGACAAACAGACATGGGTTACACAGACATGACACACACAAACGCTTACTATACACA 2913
Oy 2896 TCTACATCAGATCAAAATTAACATTTGACATCAACGAGCGCATGCAAAACCAACGATGT 2955
Db 2914 TCTACATCAGATCAAAATTAACATTTGACATCAACGAGCGCATGCAAAACCAACGATGT 2973
Oy 2956 ACGACAGAGAGGAAGATGATGAGAGAGACGTGAAACCGAGTGAAGGCGTGAAGATGAGT 3015
Db 2974 ACGACAGG-----GATGATGCAAGAGACGTGAAACCGAGTGAAGGCGTGAAGGAGTGAAGC 3027
Oy 3016 GGGTGAATGTATGAGAGGGGGGTATAGTATGATGCAATGTTATTTGTTTCATGATTTAG 3072
Db 3028 GGGTGAATGTATGAGAGGGGGGTATAGTATGATGCAATGTTATTTGTTTCATGATTTAG 3084

RESULT 4
US-10-654-416-5
; Sequence 5, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT FILING DATE: 1999-08-17
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3090
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3090)
; US-10-654-416-5

Query Match 61.9%; Score 1900.6; DB 19; Length 3090;
Beet Local Similarity 77.8%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 609; Indels 84; Gaps 8;

Oy 1 ATGGCGGGGGGCTCAAGCGGCGAGCAGAGGACACAGAAATAGCATTTGATGAGGAACAT 60
Db 1 ATGGCGGGGGGCTCAAGCGGCGGCGGCAAAAGTGCACAGAAATAGCATTTGATGAGGAACAT 60

```



Oy	61	GTTTTAGCTTTAATCTTAAGAGAGTGCCTAAGTGCAGGAATGCATAAAAACTA	120
Db	61	GTTTTAGCTTTAATCTTAAGAGAGTGCCTAAGTGCAGGAATGCATAAAAACTA	120
Oy	121	AAAAAATTTGCGCAAGATTGCTAGACCAAACTTAATATTAGAACAGTACAGAAA	180
Db	121	GAAGAAATTTCGCAAAACATTAAACAATTCAGGATTTAAATCCGAAAAGTTTACGAAAA	180
Oy	181	CTTAAAGTTTTCGGAAGATGAAAAGCAGATACAAATTCAGAACTGAAGCCCAAT	240
Db	181	TTAAAGATTTCTGTGATACGGGAAAAGAAATGAAAATGTCAGAGACTTAAAAACAA	240
Oy	241	ATTGAGAAAAAATGTACTACATCAAGAAAACTTAAGAAAGCAATTAATAAAAAATT	300
Db	241	GTCAATCAAAAATGATTAATAATTCAGAAAACTTCAAGAGCTGCAAAAAAAATT	300
Oy	301	CAGATTATTAACGGAATTAAGATTTCGCAAGAGATGAAACAATGCTCTTTTTCGAGGA	360
Db	301	TCAGAAATTAAACGATGAGATTTCGCAAAAAGATGAACAACATGCTCTTTTTCGAGGA	360
Oy	361	GTATGTTCAAAAAGACTTAAGATGATTGCATTAATTTGAGAAATTAAGTCTATCAAA	420
Db	361	GCATGTCCAAAGAACTTAAGATGACTGCATTAATTAAGAAATTAATGTTATCAAAA	420
Oy	421	AAAGCTGATAAAGTTGCGGAGAAATTTCTTTAAGAGACTTCGATGCGATCTTAATGGA	480
Db	421	GAACGGAACTAATGTGGCAGAAAGATTCTTTGAGGAGCTTCGTGTGATCTCAATGA	480
Oy	481	TCAGTCATGTGAAAAAATACTTAAGAGATTGGCCCTGTCATGCGGAGGAAAGTAT	540
Db	481	ACAAAGACATGTGAAAAAAGCTGAAAGAAATTTGCCGAAATTGAAGAGAAAGCAT	540
Oy	541	GAGTTAACAACTTGTGTCTGAACCGAAGAGACATTAAGAAATTTTAATTGAAAA	600
Db	541	GAAATTAACGAGCTTTGTCTTATCAAAAAACAATGCTGAATGTTGTAACAAAAAG	600
Oy	601	GATAAGAGTCGCGATCTTAAACAGATTTTCAGAGACTAGAAAGTTTAAAA---	657
Db	601	AAAGTTAATGTGATCTTTGAAAAAGAGTTGAAGAGCTTAAGAGAAATGAATTTG	660
Oy	658	AAAGAAATGTCCTTAATTACTCGAACATGCTATTTTACATTGGAATTTGCGAG--	715
Db	661	CGAGAAAATGCTACTTACTTAATTGAGCAATTTTCAAGAGGAACTGTGAAGGA	720
Oy	716	-----ACGACGATTAATTAATGT	735
Db	721	GACAAATCAAGTGCATTAACCTAATATAAGACTCGCAAGATATATGTACAGAGTGT	780
Oy	736	ATTGAATTTGGAGGAAATATGCCAAGAACAAATATTGCTTAATTATGCGACCGAGCCGAT	795
Db	781	GATGATTTAGCAAAAGATGTGAAAAGAAATATGTTATATGATCTCAGAGATCCAT	840
Oy	796	TTTGATCCAACTAGGCCCAGAGCTTCAATAGCAGAGATATAGGCTCGAAGAGTTTAT	855
Db	841	TTGCATCCAACTAAGCCAGAGCTTCACTAGAGAGGACATAGGCTCGAAGAGCTTTAT	900
Oy	856	AAAGAAATGTAGAGAGATGAGATTTTTAATTGAAAGAACTCATTAAGAGATGCGACGCT	915
Db	901	AAAGGGGAGAAAGAGATGGAATTTTGTGAAACAACATGTAAAGATGCAACAGCT	960
Oy	916	TTTGATTCATTTGATCCAAAGATTCTAGTCTTAATAAAAAAGACGACAAAGAGAAATGC	975
Db	961	TTGTTGGCACTACT-----TCTTAAGAAAACCTTAATAAAAAAGAAATGT	1005
Oy	976	GAAGAAAGCTTCAAAAAAGCTGCAAAATCCTCATGAACATAGAGCTTTAGAAAGTTTA	1035
Db	1006	ATAAAGCCCTTAATAAAAACTGCGAAACCTCATGAACATAGAGCCCTTAGAAAACTTA	1065
Oy	1036	TGTAAAGAAAATGTTTAATGAATGAATGAAAGAAAAATGTAAGAAATTGCAAAATGAT	1095
Db	1066	TGTAAAGAAAATTAACCAAGTATGATGAACGAAAAAATGTGATGAACCTAGAAAAAGAT	1125

OY	1096	ATTAAACAAA	CTTGCAAAATTTTTCAC	TTCCAAAGTCACTAAATATGCGCTTTTGATCCA	1155
Db	1126	GTTAAACAAA	CTTGTAACAGCTTACATCA	CAATTCCTTAAACCGCTTTACATTTC	1185
OY	1156	ACAAAGAAAT	ATATATGAAATTTGTGAT	TGGAAAGGTTGCGCAACATTTCTTAGCAAGAA	1215
Db	1186	CC-----	TGATGAAATTTGGGAAT	TGGGGAATTTACCGACATTTCTTAGTGAAGA	1236
OY	1216	GATTGTGCGAA	ATTTGGAGTCCCTATTTGTTCTATTTTGAAAAAAAATGTCAGATGAGAA	1275	
Db	1237	GATTGTGCGAA	ATCAGAAATCTTATTTGCTTTTATATTAAGAAACCTGTCCAGATGTCAA	1296	
OY	1276	AATGCATGTAAAA	TATTAAGACCAACTGTTTACAAAGAGACTTGATGCACGGGCAAT	1335	
Db	1297	GAACCTTGTAAT	TGAATGAGGGCAGCGTGTTATTAAGAGGGCTTGATGACGGGCAAC	1356	
OY	1336	AAAGTGTGCGA	AAAAATATGCGGAATGTTACATGCTCAACAAAGCTGGCTGAA	1395	
Db	1357	AGTGTGTTGCGA	AAAAATATGCGAGGTTATTTGCATGGCTCMAATTAAGATTTGGCTTAAG	1416	
OY	1396	AAAGTTTCAAC	AGAAATTAAGTAAAGTATGTAGAAACTGAAAAAAGAAAA	CAAGGAAGT 1455	
Db	1417	AAATTTCAAC	AGAAATTAAGCAAAAGTATGTAGAAACTG---	AAAGAAATTAAGGAAGT 1473	
OY	1456	TTCTCAAA	CCGATGATTTATTTATTTCTGTGTGTACGCCAGCAAAAGACCCGGTGTCTT	1515	
Db	1474	TTCTCGAAG	CGATGATTTGTTGTTCTGTGTATACACACAGCAAAAGGACGACGATTTACTT	1533	
OY	1516	ACACATATCTG	GAATGAAGAAACTATCTTTTTCGACAAACCTGATCTCAAAAGCGAGAT	1575	
Db	1534	ACACATCA	CCATCAAAATGAGAGTTATCTTTTACGACAAACCTGATCTCAAAAGAGAT	1593	
OY	1576	TTCCCGA	CAGATAAAAATTTGCAAGAAATTTGGGAGAAAGTGCCAAAGATTTTAGAGAGAT	1635	
Db	1594	TTTCCGAC	AGATTAAGACTCGAAGGATTTAGGGAGAAATTTGCCAAGATTTTAGAAGAGAT	1653	
OY	1636	TCAAAAGAAAT	TATCATGCGCATGTCTATACACTGAGACAGCAATGCATTCGTTGGGACT	1695	
Db	1654	TCAAAAGAAAT	TATCATGCGCATGTCTATACACTGAGACAGCAATGCATTCGTTAGGACTT	1713	
OY	1696	ACAAATATTTTAA	AGCGTTTTATTTGAGAAACACAAAGATCTTTGAAAGACCAAGAA	1755	
Db	1714	ACAAATATTTTAA	ACAGATTTTATTTGAGAAACAAAGATCTTTGAAAGTCAAGAA	1773	
OY	1756	AGTGTGTAAAA	ATCTATAAAGAAAGATGTAAATATGTGCTAGAAAGAGAGATGACGT	1815	
Db	1774	AACGTGTCAAA	ATTTTAAAGAAATGCCATTAATGTGCTAGAAAGGGGTGATGATCGT	1833	
OY	1816	TTCTCTTTTGT	ATGTGTCTTCCAAACGCTACGTGTGAGCTGATGTAAAGACGTGAA	1875	
Db	1834	TTTTCTTTTGT	ATGTGTCTTCCAAACGCTACAGTGTGAGCTGATGTAAAGACGTGCAA	1893	
OY	1876	GACGCGTGTAG	ATTTTCAAAAAAATATTAAGACTTCATATATTTAGAAATTTCTTGAA	1935	
Db	1894	GATGTGTGCAAA	ATTTTCGAAGAAATATGCMAGCATCAAGATATTAATGATTTCCCTTAAA	1953	
OY	1936	AATATATCAAA	TAAATATTAACAACCTGGAAGAAATTTGCGCTTTGGCATACGATATTCG	1995	
Db	1954	AAAAATCAAA	TAAAGCAATTCAGCGACGAATTTTGTCCCTCATGCGATCCATACTCG	2013	
OY	1996	AATGATTTTCA	CCCTAATGTGTCAGGTCTTACGAAGAGAAATAGT---	TGTACAAAATC 2052	
Db	2014	GATGATTTTCA	CCCAATGTGTCGATCTTAAAGAAAGAAACCTTTCTGTCAAAAATCTT	2073	
OY	2053	AAGAAGCATGT	GAGCCGTTCTATTAAGAAAGAGCCCTTGGAAAGTGTCTTCAAAAGTAGAG	2112	
Db	2074	AAAAAATAT	TATGCGAACATTTCTACAAAAGAAAGGTTTAAAGATGTCTTTAAAGTAGAG	2133	
OY	2113	CTTCAAGAAA	ATTTGACGTATTAATCTTAAATGTGAACCTGCATTTGAAAGATATTTGTACA	2172	
Db	2134	CTTGAAGAAA	TTTTAAGTATATTAACATAATTTGTAACCTGTGATTTAGAAAGATATTTGTACA	2193	
OY	2173	GTACCGGAAA	CGTAAATTAATGCGCTCAATCAGTGCGCTTATGCAAGCTAACCAAGAT	2232	

```
Db 2194 GTATTGAAAGACGTAAATATATCGTCATACACAGCTTTATGTAAAGATATATCCAAAAGT 2253
Oy 2233 AACTCTGGAAG---AGTGATGAGATGCTAGAAAAGAACTCTGTAGAAATTAAGTGA 2289
Db 2254 AAAAATTAAAAAGCCGATATATAAAAATGTTAATAAAGAACTTGTCTAATAATTAAGTGA 2313
Oy 2290 GAAATGGAAGAACAGTGCAGAACATTTACCAACAGAAATTAAGCAACCCGACAGCTGATCTA 2349
Db 2314 GAGGTGGAACAGCAATGCAAAATATTAACCAAGATTTAAACAGACTGGAATAAAGTCTA 2373
Oy 2350 AAAAAAGATTATTAACATATATAGAACTTAAGAAACGTGACAGAGAAACATTAAGCAAG 2409
Db 2374 AAAAAAGATTATTAACATATATAGAACTTAAGAAACGTGACAGAGAAACATTAAGCAAG 2433
Oy 2410 TCCAGCTTTGTTTGTCTCACTCAATTAAGAAAAGAAAGTATGTTCAAAAAGTAAATGC 2469
Db 2434 TCCAGCTTTGTTTGTCTCACTTTTAAGAAAAGAAAGTATGTTCAAAAAGTAAATGC 2493
Oy 2470 AAAAAAGATTATTAAGATCCGTTTCAACCGAGCTTCAAGATATACCAAAACATGTGAAA 2529
Db 2494 AAAAAAGATTATTAAGATCCGTTTCAACCGAGCTTCAAGATATACCAAAACATGTGAAA 2553
Oy 2530 ATACTACGAGAGGGGATTAAGATGTAATCCGTAACAGAAATTAAGATTAAGCAATTTGAT 2589
Db 2554 ATACTACGAGAGGGGATTAAGATGTAATCCGTAACAGAAATTAAGATTAAGCAATTTGAT 2613
Oy 2590 TTGGACGAGAAAGTATTTGGAAGATATGATATTTGAAGAAAGTATTAATTAATTTGAA 2649
Db 2614 TTGGACGAGAAAGTATTTGGAAGATATGATATTTGAAGAAAGTATTAATTAATTTGAA 2673
Oy 2650 TCAGATTCAGAAATTAAGAGGATTCGAAAGACTTAAGAAAGTATTCGAAAGAAATTAAT 2709
Db 2674 TCAGATTCGAGGATTAAGAGGATTCGAAAGACTTAAGAAAGTATTCGAAAGAAATTAAT 2733
Oy 2710 AAGGCTTGTGGAATCTGAAGCTCTGAGGAGTGAAGCCGCAAGAAACAGTGAACAAGT 2769
Db 2734 AAGGCTTGTGGAATCTGAAGCTCTGAGGAGTGAAGCCGCAAGAAACAGTGAACAAGT 2793
Oy 2770 ACAAGCAACAATCTAACAACAACAACAACCGTTGCCATCCGAAAGCAACGGAATGCAAA 2829
Db 2794 ACAAGCAACAATCTAACAACAACAACAACCGTTGCCATCCGAAAGCAACGGAATGCAAA 2853
Oy 2830 TCCTTACAGACAACAGACATAGGTTTACAGACATGACACACACACGCTCTACT 2889
Db 2854 TCCTTACAGACAACAGATACATGAGTTTACAGACATGACACACACACGCTCTACT 2913
Oy 2890 ATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2949
Db 2914 ATCAATCTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2973
Oy 2950 AAGTGTACAGACAGAGAGAAAGATATGACAGAGACGTGAAACCGAGTGAAGGGCTGAGG 3009
Db 2974 AAGTGTACAGACAGG-----GATGAAAGAGAGACGTGAAACCGAGTGAAGGGATTTGAAG 3027
Oy 3010 ATGAGTGGGTGGAATGTGTGAGGGGGGTGATAGTAGCAATGTTTCTTCAATGAT 3069
Db 3028 ATGAGTGGGTGGAAGCGTGTATGAGGGGGGTGATAGTAGCAATGTTTCTTCAATGAT 3087
Oy 3070 TAG 3072
Db 3088 TAG 3090
```

```
RESULT 5
US-10-654-416-9
; Sequence 9, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; protein of the human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
```

```
; CURRENT APPLICATION NUMBER: US/10/654,416
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3030)
US-10-654-416-9

Query Match 52.9%; Score 1624.8; DB 19; Length 3081;
Best Local Similarity 72.8%; Pred. No. 2.9e-311;
Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

Oy 1 ATGGCGCGGGGCTCAAGCCGCGCAG-----GCAGAGGGAGACAGAAATAGCATTTGAT 51
Db 1 ATGGCGCGGGGCTCAAGCCGCGCAGCTGCAAAAGCATCAGGGGCTAGTATATGATGAT 60
Oy 52 GAGGAACATGTTTATAGCTTTAATTTAAAGAGATGAGCTAAGTGAAGCAAAATGCAAA 111
Db 61 GAGGAATTTCTTTGGCTTTAATTTGCA---CGAATAAATTAATGATTAATGATGCAAA 117
Oy 112 AAAAAAATAAATAATTTGCAAGAAATTTGCAAGCAAAATTAATTAATTAATTAATTA 171
Db 118 AAAAAAATAAATAATTTGCAAGAAATTTGCAAGCAAAATTAATTAATTAATTAATTA 177
Oy 172 CACGAGAACTTAAGCTTTTGGAGAGTGAAGAGAGTATCAAAATGAAAGAACTG 231
Db 178 CACGAGAACTTAAGCTTTTGGAGAGTGAAGAGAGTATCAAAATGAAAGAACTG 237
Oy 232 AAACCATATTTGAGAAATAATTTGCAATCAATCAAAAGAAATTAAGAACTTAATTA 291
Db 238 AAAGAAATCTTACATCAAAATTTGCAATCAATCAAAAGAAATTAAGAACTTAATTA 297
Oy 292 AAAAAAATTCAGATTTAAGGATTAAGGATTTGCAAGAAATTAAGAACTTAATTAATTA 351
Db 298 AAAAAAATTCAGATTTAAGGATTAAGGATTTGCAAGAAATTAAGAACTTAATTAATTA 357
Oy 352 TTGAGAGGATTAATTTGCAAAAGCTTAAGATTAAGTGAATTTGAGAAATTAAGTGC 411
Db 358 TTGAGAGGATTAATTTGCAAAAGCTTAAGATTAAGTGAATTTGAGAAATTAAGTGC 417
Oy 412 TATCAAAAGAAATCTGATTAAGTTGCGGAAAGATTTCTTTAAGAGCACTTGTAGCAT 471
Db 418 TATCAAAAGAAATCTGATTAAGTTGCGGAAAGATTTCTTTAAGAGCACTTGTAGCAT 477
Oy 472 CTTAATGATCAGTATATGTAAGAAATAAATAAAGATTTGCTGTATGAGGAGG 531
Db 478 CTTAATGATCAGTATATGTAAGAAATAAATAAAGATTTGCTGTATGAGGAGG 537
Oy 532 GAAAGTGAAGTATTAAGAACTTGTGTGCTGAACAGAAAGAGATTAAGAAATTAATTTA 591
Db 538 GAAAGTGAAGTATTAAGAACTTGTGTGCTGAACAGAAAGAGATTAAGAAATTAATTTA 597
Oy 592 ATTGAAAGATTAAGAGTGCAGTATCTTTAAACAGATTTTGAAGAGCACTAG----- 646
Db 598 TTGAAAGAAATTAAGATTAAGTCTTTTAAACAGATTTTGAAGAGCACTAGAGAG 657
Oy 647 -GAAGTTTAAAGAAACATGCTTGAATTAATCTGAACATGATTTTAACTTGA 705
Db 658 AAAGTGAAGTATTAAGAGAAATTTTAACTGCTTGAACATGATTTTAACTTGAAGAG 717
Oy 706 AATTGCGAG-----ACGAGATTAATTT 729
```

D 718 AATGTGAGATATATCAAAATGTATTAATCATCCGAGACTGTATGATATTTGCCA 777  
Q 730 AAATGTATGTAATGGAGAGAAAATGCCAGACAAAATATGCTTATATGCCACGAGA 789  
D 778 GTGTGATATCATTTGGCAGTGAATAATGTGAAGAAAATAGATTATTTATACATCCGGGA 837  
Q 790 CCCGATTTTTCACATGAGCCAGAGGCTCAATAGCAGAGATATAGGCTGGAAG 849  
D 838 TCCGATTTTCATCAATCAATAGCTCAAGCTTCTAGCAGAGACATAGACTGGAAAG 897  
Q 850 TTTTATTAAGAGTAGAGAGATGAGTTTATTTGAAAGAAATCATTAAGATCCG 909  
D 898 CTTTATTAAGAGAGAGAGAGAGTTTATTTGAAAGAAATGCTTCTGTAAGAGATCA 957  
Q 910 AAGCTTTTGTGAGATGTTGATCCAGATTTCTAGCTTAAATAAAGACGAAAGAG 969  
D 958 ACTGCTTACTGGGCTTTGTGATCAAAATCTAGATCTTAAGATCAAGTGGTAAAG-- 1015  
Q 970 AAATGCGAGAGAGCCCTTCAAAAAGCTGCAAAAATCCTCATGAACATGAGGCTTAGAA 1029  
D 1016 -AATGCGAAAAAGTTCTTAAAGATATCTTAAGAGTTAAAGTCAATGAATTTTGGGA 1074  
Q 1030 AGTTTATGAAGAAAAATGTTTAAATGATGAGAAAGAAAAATGTAAGAAATGCA 1089  
D 1075 GATTTTGTATCAAAATGTAGCTGTCAAAATGAAATGAAAGTGAAGATTAAG 1134  
Q 1090 AATGATATTAACAAACTTGCAAAATTTTCTCAAAAGCTATATATCGCTTTT 1149  
D 1135 AAGGAGTTAGCAACAGTATCTAAATATCTTTTAAATAAAGAAATTAACACTCTCT 1194  
Q 1150 GATCAACAAAAGGAAATGAATGTAATGTTGGAGGAGGAGGCCAACTTTCTTAGC 1209  
D 1195 -----GATCCGAGAGAGTCAATTCATGATATGATGAGACATTTCTTAGT 1242  
Q 1210 AACGAGATGTGTGGAATGTGAGTCTTATTTTCTAATTTGAAAAAATGTCAGAT 1269  
D 1243 GACATATCTGCAACAGTTTGAAGTCTGAGATGTTTATTTT---AAAAGTCAAGCACT 1299  
Q 1270 GGAGAAATGATGTTAAATATTAAGCAATGTTTACAAAGAGACTTGATGACCG 1329  
D 1300 CTGACAAAGAGTGTATTAATCTAAGGCAATGTTTAAAGAGGCTTGAAGCAAA 1359  
Q 1330 GCAATTAAGTCTGCAAGAAAATATGCAAGAAATGTTTCAATGTTTCAAAAGCTGG 1389  
D 1360 GCTAATAGAGATTCGAGAAAAGATGACGACTGTTCTATGTTTCAAGGAAAGATGG 1419  
Q 1390 CTTGAAAAAGTTTCAACAGAAATAGTAAAGATGAGAACTGAATAAAGAAACAA 1449  
D 1420 TTTTAAAGAACTATAGAAAAAATATGAGAAATGTTGGAATCTTAAACA----- 1470  
Q 1450 GGAAGTTTCTCAACGATGATTTATTTATTTGTGTGTACAGCCGCAAAAAGCCCG 1509  
D 1471 -----ACAAAGTGAAGTGTGTTTGTCTATGATATGATCACTTAAAGCATGAGA 1521  
Q 1510 TTGCTTACATGATCTTTCGAATGAAAACTATCTTTTACGACAACTGATCAAAAG 1569  
D 1522 ATACTTTCAGCTGATATCTCAAGCAAGCACTCTTTTTCGAAAAATTTGATCAAAAG 1581  
Q 1570 CGAGATTTCCGACAGATTAATAATGCAAGAAATGGGAGAAAAGTCCAGATTTTGA 1629  
D 1582 CGAATCTTCCAAACAGAAAGATTTGCAAGAAATTAAGAAAGAAATGTAAGCTTTTGG 1641  
Q 1630 GAGATTTCAAAAGAAATTAATCATGAGCTATATCATGAGAGCAATGCAATGCTTG 1689  
D 1642 AAGGATTCAAATCAGATTAAGTGGCCATGTCATACGCTTAAACAAAGTGTGATGCTTG 1701  
Q 1690 GGAAGCTACAGAAATTTTAAAGCAGGTTTATTTGATGAAACAAAGATTAATTTGAAGAC 1749  
D 1702 GGGATCTACAGAAATCTTGAACAGGTTTATCTAGATGAACAAAGATATCTTTAAGAACT 1761  
Q 1750 CAAGAAAGTTGTATAATATCTTAAAGAAAGATGTAATAATGCTTGAAGAGAGAT 1809  
D 1762 CATGAAAACTGTACGAATAATTTAAAGAAATGTATTAATGCTTGAAGAGGAGTAT 1821

Q 1810 GACCGTTTCTTTTGTATGTCTTCCAAAAGCTACGTGTGAGCTGATGTAAGAAC 1869  
D 1822 GATCGTTTCTTTTGTATGTGTTTACCAAAAGCTACGTGATGATGATGAT 1881  
Q 1870 GTGAAGACAGGTGTGAAGTATCAAAAAAATTAATAAGCTTCAATATTAATTTGAATTT 1929  
D 1882 GTGAAGACAGGTGTGAAGTATTTAAAAAATATGCAACGCTCAGATTAATTAATTTCT 1941  
Q 1930 CTTGAAAAATTAATAATAATAACAACTGGAAGAAATTTGCTTCCCTTGGCATAGC 1989  
D 1942 CTTAAAAATTAACAAATTAATAAAGAAATCAGCAAAATATTTGTCTTATGACCA 2001  
Q 1990 TATTCATATGATTTTCACTAATGTCAGTCTTTCAGAAAGATAGT---TGTACA 2046  
D 2002 TACTGCGATGATTTTATCCCAATGTCTGATCTTAAAGAAAGAAAACTTCTGTGCA 2061  
Q 2047 AAAATCAAGAACATGTCAGCGGTTCTATPAAGAAAGGCTTGAAGATGCTCTCAA 2106  
D 2062 AATCTTAAAAATATGCGAAACATCTCAAAAAGAGGTTTGAAGATGCTCTTAAA 2121  
Q 2107 GTAGAGCTTCAAGAAAAATGACTGATTAATCTTAAATGTAACCTGATGAAGAAATAT 2166  
D 2122 GTAGAGCTTCAAGGAAATTAAGTAAATGAATGAATGAATCTGATTAAGAAAGATAT 2181  
Q 2167 TGTACAGTACGGGAAACGTAAATATGCGTCAATCAGTGGCTTAATGCAAAAGCTAACAC 2226  
D 2182 TGCATATATTTGAATAATGTAAGTATTCATCAATCAAGTTTATTAATGAATTAATCC 2241  
Q 2227 AAGGATTAATCTGGAAGAGTATGAGATGCTAGAAAGAACTGTGAGAAATTAATG 2286  
D 2242 GAAAGTAAATCTAAAAAGACGATTAATGAAGTATGAAGAAAGCTTTGTCTTAAATTAATG 2301  
Q 2287 AAAGAGTGAAGAACGTGCAAAAGCTTACCAAGAAATTAAGACAAACCGGACGCTGAT 2346  
D 2302 GAAGAGTGAAGAACGATTAATTAATGTTACAGCAAAATTTGAGCATGAGGAAAAAGAC 2361  
Q 2347 CTTAAAAAAGATTAAGACATATGAGAACTTAAAGAAAGTGCAGAGAGAAAGCAATGAC 2406  
D 2362 CTTAAAGATGATTTTGAACATTTTGAATAAATCTTAAATAACAGCAGAGAAACATGAT 2421  
Q 2407 AAGTCAGTCTTGTGTTTGTCACTCATTAAGAAAAAGAAATTAATGATCAAAAAATGAT 2466  
D 2422 AAATCCATCTTGTGTTTATCATTCGTTAAGAAAGATGAATAATATATCATGAAAAATGAT 2481  
Q 2467 AGCAAAAACAGATTAAGATGCGGTTTCAAGGAACTTCAAGATACCAAAAAATGATG 2526  
D 2482 AGCAAAAGCAAGATTAAGAAATCGTTTCAAGGAACTTCAAGATACCAAGAAATGATG 2541  
Q 2527 AAAATCTAGGAGGAGTTAAGAGATGATCCGTATCAGAAATTAAGAAAGCTTAAAGATTT 2586  
D 2542 AAAATCTAGGAGGAGTTAAGAGATGATCCGTATCAGAAATTAAGAAAGCTTAAAGATTT 2601  
Q 2587 GATTTGGCAGCAGAAATTAATGGAAGATGATGATGATTTGAAGAAAGATGATTAATTTG 2646  
D 2602 GATTTGGTACAGAAATTTGGAAGATCTGACTTGAAGAAAGATGATTAATTTG 2661  
Q 2647 GAATCAGATTCAGAAATTAAGAGATTTGCAAGAACTTAAAGAAAGATGCAAAAAATTT 2706  
D 2662 GAATCAGATTTGAGAGATTAAGAGATTTGCAAGAAATTAAGAAAGATGCAAAAAATTT 2721  
Q 2707 AATAAGCTTGTGCAATGGAAGCTCTGAGAGTGAAGCGCAGAAACATGATGACAGAA 2766  
D 2722 CAAGAGTATGTTTGAATTAATTAACCTGAAGATGAAGCGCAGAAACATGATGACAGAA 2781  
Q 2767 AGTACAAAGCACTACAAACAAACAAACCGTTTCCGATCCGAAAGCAACGAAATGC 2826  
D 2782 AGCAAAAGCAAGCAAGAGCAACAAACCGTTTCTGATCCGAAAGCAACGAAATGC 2841  
Q 2827 AAATCTTACAGACACAGACATGAGTTTACAGACATGACACACACAGACGCTCT 2886  
D 2842 AAATCTTACAGACACAGATATGATGATTAACAGACTTGCACACATACACAGACGCTCT 2901

Qy	2887	ACTATCACTATTACCATTACAGTTCAAAAATAACATTGACATCAAGAGCGCATGGAAACA	2946
Db	2902	ACCATTCACATTACATCATTACATCAAAAATAAACCCTCACATTAACCAAGCGGTGGAAACA	2961
Qy	2947	ACCAAAGTGTAGCAGCAGAGGAAGAATGATGCAGAGACGTGA AACCCAGTGAAGGCGTG	3006
Db	2962	ACCAAAGTGTACGACAG-----GGATGATGCAAGAGACGTGA AACCCAGTGAAGGATTG	3015
Qy	3007	AGAGTAGAGTGGTGGAAATGTGATGATGAGGGGGTGTAGTAGCAATGTTATTTCTTCATG	3066
Db	3016	AAGATGATGGGTGA AAGTGTATGATGAGGGGGTGTAGTAGCAATGTTATTTCTTCATG	3075
Qy	3067	ATTTAG 3072	
Db	3076	ATTTAG 3081	

```

RESULT 6
US-10-654-416-1
/ Sequence 1, Application US/10654416
/ Publication No. US20040043409A1
/ GENERAL INFORMATION:
/   APPLICANT: Kovacs, et al.
/   TITLE OF INVENTION: Identification of a region of the major surface
/   TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
/   FILE REFERENCE: 4239-58054
/   CURRENT APPLICATION NUMBER: US/10/654,416
/   CURRENT FILING DATE: 2003-09-02
/   PRIOR APPLICATION NUMBER: US/09/762,724
/   PRIOR FILING DATE: 2001-02-09
/   PRIOR APPLICATION NUMBER: PCT/US99/18750
/   PRIOR FILING DATE: 1999-08-17
/   PRIOR APPLICATION NUMBER: US 60/096,805
/   PRIOR FILING DATE: 1998-08-17
/   NUMBER OF SEQ ID NOS: 26
/   SOFTWARE: PatentIn Ver. 2.0
/   SEQ ID NO 1
/   LENGTH: 3042
/   TYPE: DNA
/   ORGANISM: Pneumocystis carinii sp. f. hominis
/   FEATURE:
/     NAME/KEY: CDS
/     LOCATION: (1)..(3042)
/     US-10-654-416-1

```

Query Match	35.8%;	Score 1098.6;	DB 19;	Length 3042;
Best Local Similarity	63.0%;	Pred. No. 4.5e-207;		
Matches 1954; Conservative	0;	Mismatches 1954;	Indels 93;	Gaps 13;

Oy	2	TGGCGCGGGCCGGTCAACGGCGGACGGACGACACAGAAATACATTGATGGAGAA---C	58
Db	2	TGGCGCGGGCCGGTAAAGCGGACGGTACCGAGACATCAGAGTAGATGAGAGAGAAATGGC	61
Oy	59	ATGTTTATAGCTTAATTTCTAAAGAGATGACTAAGTAGCAGCAATGCCAAAAAAC	118
Db	62	GCTTTTGGCTTAATTAATCAAAAGAAATTTCTAAGAGTAGTAAAAATGGAAGAAAAAT	121
Oy	119	TAAAAAATATTGGCCAGATTGACTGAAAGCAAATCTAAATATGAAACAATGCACAGAA	178
Db	122	TAGAAAAACATTGCAAAGATTAAAGTGAAGCAATCTAACCTCCAGAACAAAGTACATGAAA	181
Oy	179	AACTTAAAGGTTTGGCGAGATGCGAAAAAGACGATCAAAATGCAAAAGAACTGAAGCCG	238
Db	182	AGTTAAAAAGATTTCTGTGATAGCAAAAAACGTGATAAAAATTTAAAGAACTTAAAAAAA	241
Oy	239	ATATTGAGAAAAATGTACTACATCACTAAAGAAAACTTAAAGAAAGCAATTTAAAAA	298
Db	242	ATGTTGAAAAAAATAGCGGTGATTTTAAAAACAGAAATTGAGAAATTTGTGAAAAAGAG	301
Oy	299	TTCAATTATTAACGATAGAGATTGCAAAAGAAATGAACAAATGCCATTTTGTGAAG	358
Db	302	CTTCAAAATTTGAAAAATGATGAGTGCACAAAAAATGAACAACAGTCTTGTTTGTGAAG	361

QY	35	GAGTATCTGCAAAAAGAACTTAAAGATGATGGCAATCTTTGAGAAAATTAAGTCATCAAA	418
Db	362	AAGATGCTC--TGATCTTACAAAGATGGCAACGATTTAAGAACAAATGTTATCAGA	418
QY	419	AGAAACGTGATTAAGTTGGCGAAGAAAGTTCTTTTAAGAGCACT--TCGTAGCATCTTA	475
Db	419	ATAAGCGTGAATAGGTAGCAAGAAAGACTTTTAAAGATTAATTAAGAAAGAAAGATTTTA	478
QY	476	ATGATCAGTCATATGTGAAAAAAAACCTTAAGAGATTTGCCCTGTCATGGGAGGAA	535
Db	479	AAGATTAATAATTCATGTGAAAAATTAAGTCGAGATATCTCTCAAGAAATTAAGTCAATATGA	538
QY	536	GTGATGAGCTTAACAACTGTGTCTGAACGAAAGAGACATGTAAGATTAATTTTAATG	595
Db	539	GTGACGAATTAAGTAAATTAATGTTTTGATCAAAAAATACGTGTGATTAATCTTGTAAMG	598
QY	596	AAAAAGATTAAGAAGTG-----CGGTACTCTTAATAACAGATGTTTCAGACACATAG	646
Db	599	AAACGCCAACAAAGGTGGAATCTTTCAAAAATCTTAAACCGAAATTTAAACATTAAGG	658
QY	647	GAAGTTTTAAAAAAGAAACATGTCTTGAATTACTCGAACAATGCTAATTTTACATTTGAA	706
Db	659	AAGATGAACAACATAAAAAAATGCCCCATTAATATATGAAGATGCAATTTTATATGATG	718
QY	707	ATTGCGGAGACGAGATATATTAATATGTAATTGAATTTGGAGGAAAAATGCCAAGACAA	766
Db	719	AAAGTTGTGGAAACGATTCACCTGAAGTGTAGTGAATTTGAAAAAATATGTCAGAGAAAA	778
QY	767	ATATTGCTTAATATGCCACGAGAACCCGATTTTGAATCCAACTAGGCCGAGGCTACATATAG	826
Db	779	ATATTACTTACACATTAATTCATATTCAGGGTTTGATCTTAATAGAACCGAAATTAATCATTAG	838
QY	827	CAGAGATATATAGGCTGGAAGAGTTTATTAAGAAGTGAAGAGATGAGATGATTTTATATG	886
Db	839	CAGAAAGATTAAGACTTGAAGGAATTTATGAAGAACGACGAAGAAAGAACTCTTGTTG	898
QY	887	GAAAGAAATCATCTAAGAGATGCGACAGCTTTGTTGGCATTTGTTGATTCGAAGATTTCTATGC	946
Db	899	GGAATCCCTTATACGACGATGCTACTGCTTTGGTGCAATTTTGAATTCAGAGATCCATCTC	958
QY	947	TTTAAA---AAAAAGACGACAAAGAGAAATGGCGAAGAACCCCTTCAAAAAAGCTGCATA	1003
Db	959	TTACAACTCAACGACCTACAAAGAAAAATGTAATAAAATTTCTTGAAGATTAATGTTAATA	1018
QY	1004	ATCTCATGAACATGAGGCTTTTGAAGATTTATGTATGAAGAAAAATGCTTTAAGTAATGATG	1063
Db	1019	ATTTAAGAAACATGATATTTATTAAGAGTCTATGCCAGATTTATTAATGCAATTAATACATA	1078
QY	1064	GAACGAAAAAATGTGAAGAAATGGCAAAATGATTTAAACAAACCTTGCAAAATTTTCACTT	1123
Db	1079	AGGACAAAAAATGCGAAGAACTTAGTCAATATATGGAAGAAACATGTAAAAATTTTTCATTT	1138
QY	1124	CAAAAGTCACATTAATATCGTCTTTTGAATCCAAAGAAATTAAGAAATTTGTTGAT	1183
Db	1139	CAAAAACCCCTATGATTCATTTTGTGGCATGGAATTAATAAATGATGAGAAATTAATTAAT	1198
QY	1184	GGGAAAGGTTGCCCAATCTTCTTAGCAACGAATGTGTGCAAAATTTGGAGTCTCTAATGTT	1243
Db	1199	GGGGGAATTTATCAAGCTTTCTAAGCAATTAAGATTTGTACAAATTTGAAATCGATTTGTCT	1258
QY	1244	TCATATTTGAAAAAATGTCAGATGAGAAAAATGCAATTAATAATTAAGAGCAACAT	1303
Db	1259	TTTATTTTGAATAAAGCTGTAGA--AGCGAAATCTGATCAAGAAATATATCAGACACAT	1315
QY	1304	GTTACAAAAAGAGACTTGATGACAGGGCAATAAAGTGTGCAAGAAATATATGCGAGGA	1363
Db	1316	GCTACAGAGAGGACTTGACATATTACCAATATGAATTAATTAACAAAAAGAAATGCGAGGA	1375
QY	1364	TGTTATCATGTGTTCAACAAAGAGCTGGCTGAAAAAGTTCAACAAGAAATTTGTAAGATAT	1423
Db	1376	TGCTGCAATGGTTCAATTAATCAATGCTGTTATGTTGTTTCCAAAAAATCTCATAGAAATGT	1435
QY	1424	GTGAGAACTGAAAAAGAAAAACAAGAAAGTTTCTCAACGATGATTAATTAATTTCTGT	1483

```
Db      1436 GCAAAAAAGTAAAAAGAGATTAAGAGATTTTCCGAGTAATGATTAATTTGCTTAT 1495
Qy      1484 GTGTACGCCAGCAAAAGCAGCCGGTTCCTTACACATGATCTTCGATGATAAACTATCT 1543
Db      1496 GTGTACCAACCATCAAAAGCAGCTCGATTGCTTCGATGATCTTCGATGATAAACTATCT 1555
Qy      1544 TTTTACGACAAACATCGATCAAAAGCGAGATTTCCGACAGATTAATAATTCAGAGAAAT 1603
Db      1556 TTTTGCAGACGATTTTGAACGAAAGCGAGATTTCCAGTGAAAGAAAGACTCGCAGAAAT 1615
Qy      1604 TGGGAGAAAGTCCCAAGATTTTACGAGAGATTTCAAAAGAAATTCATGGCCATGTCATA 1663
Db      1616 TATTTAAAGAAATGTGAGGCTCTAAGAAAGATCTTAAAAAAATTGAATGGCAGATGCATA 1675
Qy      1664 CACTGAGCGCAATGCATTCGCTTGGGACTACAGAAATTTTAAACAGGTTTATATGG 1723
Db      1676 CATTAAAGCCAAATTTGTATCAATTTGAGAAACGCTAAAGAAATGAAAGAACTTTTACTAA 1735
Qy      1724 ATGAAACAAGATACCTTTGAAAGACCAAGAAAGTTGTAAATTAACCTAAAGAAAGT 1783
Db      1736 ATGAAACTTAAGATATATTGAAAAATCAAGAAATTTGTGAATGTATTTGAAGAGAAAT 1795
Qy      1784 GTATTAATGCTCTAGAAGAGAGATGACGTTTCTTTTGTATGTGTCTTCCAAACG 1843
Db      1796 GCAATGAATGCTCTAGAAGAGAAATGACGTTTCTCTTTTATGTGCTTTCGAAATA 1855
Qy      1844 CTACGTGTGAGCTATGTATAAGACGTGAAGACAGGTGTGAAGTATTCAAAAAAATA 1903
Db      1856 GGCCTTGCAAGATATGTGTAGAAATGTGAAATTCATGCAAAATTTTGAAAAAACA 1915
Qy      1904 TAAAGCTTCATATATTGAATTTCTTGAAATTAATACAAATTAATTAACAACATCG 1963
Db      1916 TTAAAAAATACCAAGATTTG-----ATAGTAATCAATTAATTAAGAAATTAAG 1963
Qy      1964 AAGAAATGTCTCCTCTTGGCATACGATATGCAATGATTTTCACTTAATTTGTCAGAGTC 2023
Db      1964 GGCATATTTGCTTAATTTGGACCCACACGCGCATAGTTTGGACCCAAATTTGCCGATC 2023
Qy      2024 TTGAAAAA-----AATTAATGTGAACCTTTGAAAAATTTGCAAACTTATTAAGCAAA 2080
Db      2084 AGGCTTGAAGATGCTCTCAAGTAGAGCTCAAGAAATTAAGCTGATTAATCTAAAT 2143
Qy      2081 GAGACTTTGAAATGCACTTATATTTGATTTAGAGACATCTTGATTAAGAAAAAACT 2140
Db      2144 GTGAACTGCAATGAAAGATATTGTAACAGTACGGGAAACGTAAATTAATGCGTCAATCA 2203
Qy      2141 GCAAAACAAATCTTGATTAAGTACGTACATATGGGATCAAAACGAAATTAACACCTTA 2200
Db      2204 GTGGCTTATGCAAGCTTAACACCAAGATTAACCTGAAAGATGAGAGCTTGAA 2263
Qy      2201 AAGGTTTTTGTG-----ACAGTTCTACTGATTAACAAAGAAACATTTAGAG 2245
Db      2264 AGGAACTCTGTGAAATTAAGTGAAGAGTGAAGAAAGCTGAAGCAATTAACAACAG 2233
Qy      2246 ATAAACTTTGGAAAACTAGTTCAGGTGGAAGAAAAATGCAAGATTAACAAG 2305
Db      2324 AATTAGCAACACCGCAGCTGATCTAAAAAAAGATTAAGACATATGAGAACTTAAGA 2383
Qy      2306 AACTTGAAGAAAGCAAAATGATTTAGAGAAAAACATTAAGATTAAGAAAAATTAAGA 2365
Db      2384 AACTGTGAGAGAAAGCAATGCAAGTCCAGTCTTTGTTGCTCATCTATTAAGAAAAAG 2443
Qy      2366 AGGATTAACAAAAATGCAATGGAAGAAACAAATCTGTTTTTCAACAATTAACCAACAG 2425
Db      2444 AAGTAAATGATCAAAAGTAAATGACAAAAACAAGATTAAGAAAGCGGTTTCAACGAC 2503
Qy      2426 ATATTAATAACGAAAAAG-----AGTCAGACCTTAATGACGCTTAGTG 2470
Db      2504 TTCAAGATACCAAAACATGTGAAATACTACGAGAGGAGTTAAGAGTATACGTTAA 2563
```

```
Db      2471 TTCAGATATTGTACATTTT-----AACTGTAAAAAGAAATGAAGAAAGTCAAGTGA 2524
Qy      2564 CAGAAATTAGAGGCTAAAGCATTTGATTTGCGACAGAAAGTATTTGGAAGATATGATGATT 2623
Db      2525 CAGAAAAAGAAAGCAAAAGCCTTTGATTTGATTTGATGACATGACATTCAGCTTTATGTAAGT 2584
Qy      2624 TGAAGGAAGATGTAAATTAATTTGAAATCAGATTGCAAGAAATTAAGAGAGATTGCAAGACT 2683
Db      2585 TAAAGAAACGTGTACACATCTAAAGGATGATTCGCAATTTAGAAAAAGATGTA---T 2641
Qy      2684 TAGAAGAGTATGCAAAAAAGATTAATTAAGCTTTGCGCAATCGAAACCTCTGAGGTGA 2743
Db      2642 GTAAAGCCAGTCAAGATGAAGATGAATAATATGTTTAAATAAGAAACACTGAAGTAA 2701
Qy      2744 AGCCGACAGAAACGTGACAGAAAGTACAGACGCAACTACACAAACAACAACCGTTG 2803
Db      2702 AGCACAATGAAATTAATAACGTAACGAAACCAACATTAACAAGTACAGAAACAGTCA 2761
Qy      2804 CCGATCCGAAAGGCAACG-----AATGCAAAATCCTTAACAGACAAACAGACA 2848
Db      2762 AAGAGCAGAAAAAACAAGTAGAGAGACGAGAGAAATGCAAAATCTCTCAGACAAACAGACA 2821
Qy      2849 CATGGTTACACAGACATGACACACACACAGACGTTACTATCAATCTAACATCAAT 2908
Db      2822 CGTGGGTCACAAAAGACGTCAACCCATACACAGACCTTCACAGACTACGTCCACAGTTACGT 2881
Qy      2909 CAAAAATTAATTAATGACATCAAGAGAGGCGATCAACCAACCAAGTATACGACAGAGAG 2968
Db      2882 CAGAAATTAACATGACCTTGACAGAGGCGGTGTAAAGCTTACAAAGTGTACAGAGAGAG 2941
Qy      2969 AAGATGATGACAGAGACGTGAAACCGAGTAGGGGCTGAGAGATGAGTGGTGAATGGA 3028
Db      2942 AAGATTAAGCAGAGAGAGGTGAAGCCGAGTGAAGGGGCTGAGAGATGAGTGGTGAATGGA 3001
Qy      3029 TGAAGGGGGGTATGATGATGCAATGTTATTTGTTGATGATT 3069
Db      3002 TGAAGGGGGGTATGATGATGATGATGATTCATTCATGATT 3042
```

```
RESULT 7
US-10-654-416-3
; Sequence 3, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; Applicant: Kovacs, et al.
; TITLE OR INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patencim Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3006)
US-10-654-416-3
```

```
Query Match      30.2%; Score 927.4; DB 19; Length 3006;
Best Local Similarity 60.3%; Pred. No. 3,5e-173;
Matches 1859; Conservative 0; Mismatches 1131; Indels 93; Gaps 16;

Qy      2 TGGCGCGGGCGGTCAAGCGCAGGACAGGAGACAGAAATGACATTTGATGAGAAACATG 61
```

Db 2 TGGCCGCGGCGCTCAAGCGCGGCTGCAG--CACAGAAATGCTTGAAGAAATATC 58  
Qy 62 TTTTACCTTAATCTTAAGGAAGATGGACTAAGTACAGAAATGCAAAAAAACTAA 121  
Db 59 TTTTGGCTTTGATTTTGAATAATGATGATAATTAATGATTAATGATTAATGATTA 118  
Qy 122 AAAAAATTTGCCAAGAAATGACTGAAGCAAACTTAAATATAGAACAACTACAGAAAC 181  
Db 119 AAGAGTATTTGAAAGTTTAAAAATGTAAACAAAAGCAAAAAAACTTAAGAAAGT 178  
Qy 182 TTAAGGTTTTTGGCAAGATGAAAAAGC---AGATACAAAATGCAAGAACTGAAGCA 238  
Db 179 TAGACGGAATCTGCAAGATGATTAACAAATAGAACAAAATGCAAGAAATCAAGAA 238  
Qy 239 ATATTGAGAAAAATGTACTAACATCAAGAAAACTTAAAGAACAAATTAATAAAAA 298  
Db 239 AGGTTAAAGCAAGTACTAGTTTCAACGAACTTAAAGCACTCAAAAAAGGAG 298  
Qy 299 TTCAGATTATACGATTAAGATTTGCAAGAAATGAACCAATGCTATTTTGGAG 358  
Db 299 CTTCAACATTGAAAGATATATATGTAAAGAAAGCAACATGCTGTTTTTGGAG 358  
Qy 359 GAGTATGTTCAAAAGAACTTAAAGATGATGCAATCTTGAAGAAATAGCTATCAA 418  
Db 359 GAGCATGTCACAGAACTTAAAGTAAATGATTAATGAACTGAGAAATTAATGTTATCAA 418  
Qy 419 AGAAACGTATTAAGTTGCGGAAGAGTTCTTTTAAAGCACTTGTGACATCTTAATG 478  
Db 419 AAAAAAGAGACGATGACAAAAAAAGCTTTTAAAGTACTTAAAGGAACTTAAAG 478  
Qy 479 GATCAGTATATGTGAAAAAACTTAAAGATTTGGCTGTGCTGAGGAGGAAAGT 538  
Db 479 ATAAAAACACATGCAAAAAATTAAGGGGGTTTGTCAAGAAATTAACAAAGAAATG 538  
Qy 539 ATGAGTTACAAACTTGTGTGCAACAGAAAGACATGTAAGAAATTTTAATTGAA 598  
Db 539 ATGAGCTTAATAATTAATGTCTTGACGAAGAAAAACGTGTGAGATCTTGTATCTAAG 598  
Qy 599 AAGATTAAGAGTGCCTACTCTTAAACAGATGTTTACAGACACTAGAA-----GTT 652  
Db 599 AAGATTAACAAATGCAACCTCTCAAAAGAGAAATGATCTAGTGTGGAAGAAATG 658  
Qy 653 TTTAAAAAGAAACATGTCTGTAATTAATCTGAAACAAATGCTATTTTACATGGAATTCG 712  
Db 659 TATTAAGAAAGAAAGTTTATTTCTTGAAGAAATGTACTTTATGAGTCAACCTGTG 718  
Qy 713 GAGACGACGATATAATTAATGTATTTGATTTGGAGAAAAATGCCAAGAACAAATATTG 772  
Db 719 AAACGATCAGCA---AAGGTAAAGAGTTTGCAGCAAGAAATGTCAAAGAAATCTCG 775  
Qy 773 CTTATATGCAACAGGACCCGATTTGATCCAACTAGGCAAGAGCTAACATAGCAGAG 832  
Db 776 TTTATGACGACACAGGTTCACTTGTGATCTTAAAGAAATGATGTTAGCAGAG 835  
Qy 833 ATATAGGCTGGAAGATTTTATAGAAAGTAAAGAGAGATGAGATTTTATGGAAGA 892  
Db 836 AAATAGACCTTGAAGAAATTTGTACGTAAAGCAGTAAAAAGGAAATATATGGAAGC 895  
Qy 893 ATCATCTTAAGATGCGACAGCTTTGTGCAATTTGTATCCAAAGATCTAGCTTAATA 952  
Db 896 CATCAATTAAGATGAGTACGCTTTATTTGCAATTTTAAGCA-----GAGTATGCTC 949  
Qy 953 AAAAAAGCAGCAAGAAAGATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATG 1012  
Db 950 AAAATATCTTTTAAAGATCAATGTGAAGATGTTATTAATAAAAAATGTGAAAACCTTAAAG 1009  
Qy 1013 AACATGAGGCTTTAGAAAGTTTATAGTAAAGAAATGTTTAAAGTATGATGGAAGAA 1072  
Db 1010 AGCATATTAATTTTAAAGATTAATGATTAATTAAGCATATCACTGATATCCAAAAAGAA 1069  
Qy 1073 AATGTGAAGAAATTTGCAAAATGATATTAACAAAATTTGCAAAATTTTCACTTCAAAAGTCA 1132  
Db 1070 AATGGAAGAACTAAATTAAGAGTT-----AACCAACCGTATTTTAACTGTTTCTAATA 1123

Qy 1133 CTATTAATCGCTCTTTTGTATCTCAACAAAGAAATTAATGAATTTGTTGATGGAAGGT 1192  
Db 1124 GGAATGGAATAATTTTGTCTCAGCTAATGTAA--GGAATTAATGTTGGGATATGT 1180  
Qy 1193 TGCCACATTTCTTAAGCAAGAAATGTTGCGAAATTTGAGTCTTATGTTTCTATTTTG 1252  
Db 1181 TGCAATACATTTCTTTGGTGAAGAGAGTGTACGAACTGTGTCCGATTTGTTTTTATTTGA 1240  
Qy 1253 AAAAAAATGTCCAGATGAGAAAAATCATGTAAAAATATAGACCAATGTTACAAA 1312  
Db 1241 AAGCCAAAGCTCCACTT---GAAAGCCCTGCAATTACTTAAAGCAGCATGTTATATAA 1297  
Qy 1313 GAGCATTTGATGACCGGCAATTAATGCTGCAAGAAATATGCGAGAAATGTTACATG 1372  
Db 1298 AAGGCTTTGAAGCGTGTGCAATGAACTAACTAACTAAAGTTACGCGGAAAAATTTGCAAG 1357  
Qy 1373 GTTCAACAAAGCTGCTTGAAGAAAGTTTCAACAAAGATTAATGTAAGTATGTAGAAAC 1432  
Db 1358 GTTCAATATGAAACATGCTTGAACCTTCAAAAAAACTTGTAAAGTTTGTAAAGAA 1417  
Qy 1433 TGAAGAAAAAGAAACAAAGAAATTTCTCAACGATGAATTAATTTCTGTGTACAGC 1492  
Db 1418 CGAAAGAGAG-----AAGTATGAATTAATTTGTACTATGATGAACC 1459  
Qy 1493 CAGCAAAAGCAGCCGCTGCTTACATATGATCTTCCAAATGAATACTATCTTTTACAGC 1552  
Db 1460 CAATTAAGAGCGCTTACAGTGTCAACAGATTTGCAATGAGGAGCGATGCTTTGCAAG 1519  
Qy 1553 AACAACTGATCAAAAGCAGATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGAGAA 1612  
Db 1520 AGCATTTGAACGAAAAAGAGATTTCCACAGAAAAAGATTTGTAAGAAATTAAGAAAA 1579  
Qy 1613 AGTCCAAAGTTTGAAGAGATTCAAAGAAATTAATGATGCGCATATCACTGAGC 1672  
Db 1580 AATGTAGGCTTTAGAAAAAGATTCAGAGAAATTAATGATGATATGATTAATGAAC 1639  
Qy 1673 AGCAATGCAATCGCTTGGGACTACAGAAATTTTAAAGCAGGTTTTATGATGAACCA 1732  
Db 1640 AGCATTCGAATCGCTAAGAGCATAGACCTTAAAGAGAGAGTTGCTAAAGAAAAATA 1699  
Qy 1733 AAGATCTTTGAAGAACCAAGAAATGTTGTAAATACCTTAAAGAAAAAGTAAATAAT 1792  
Db 1700 AAGATTTTAAAGATGAATAATGCTGCAAGAAAGAGCTTAAGAAACGATGTGAATAAT 1759  
Qy 1793 GGTCTAGAGAGAGATGACCGTTCTCTTTGTATGTCTTTCAAACGCTACGTGTG 1852  
Db 1760 GTTTGAAGAGAAATTAATAATTTTTCGCTGTTGTGACTGTGAACCTTGTGCA 1819  
Qy 1853 AGCTGATGTTAAAGACGTGAAGACAGGTGTGAAGTTTCAAAAAAATTAAGCTT 1912  
Db 1820 AAAAGATCACTAAGAAATGTGAATCTAAATGTATATATGAAGAGACATGTGAACTA 1879  
Qy 1913 CATATATTATGAATTTCTGAAAAATATACAAATTAATAACAACCTGAAAGAAAT 1972  
Db 1880 TGAACGTTTATAGGAATTAAGCTTAAAAAGGAAAAAATATGTGAAT-----TT 1930  
Qy 1973 GTCCCTCTTGCAATAGATTTGCAATTAATTTGCTCAAGCTTACGAAG 2032  
Db 1931 GGGCTCAATATTTAAAGATGAAGCAAAATTTGAAAACTTAAAAACGAGAAAG 1990  
Qy 2033 AGAATAGTTGTACAAAATTAAGAGCATGCTGAGCCGTTCTATTAAGAAAGCCTTGG 2092  
Db 1991 ATGGCAATGCAAAAACTCAATTAAGAGTCAATCAATCTTGAAGAAAAAGCCTTAG 2050  
Qy 2093 AAGATGCTCTCAAGTGTAGAGTTCAAGAAATTAAGCATTAATTAATATGTAAGTGAACCTG 2152  
Db 2051 AAAATTAAGTTGTAAAGAAATTAAGAGTATTAATTAAGAGAAATGTAACAATA 2110  
Qy 2153 CATTGAAGAAATTTGTACAGTACGCGGAAAGCTTAATTAATGCTCATACGTGGCTTAT 2212  
Db 2111 CACTTAATATATACGTACCAATTTGAAGAAAGCAGAGAAATGGGTGGAACCTTGT--- 2167



2213 GCAAGCTAACCAAGATTAAGTGGAGAGTGAAGTGTAGAAAGAACTCT 2272  
2168 -----GCAAAAGGAAAGAAACAC---CAAGATGACATTAAGTTAGAGAACTCT 2218  
2273 GTGGAATTAATGTAAGAAAGTGAAGAACTGCAAGCTTACCAAGAAATTAGAC 2332  
2219 GTGAAAGCTAAATTAAGTAAAGAAATCTCAAAATTGAGAGACGCTTGAG 2278  
2333 AACGGAGCTGATCTAAAGAAATTAAGACATTAAGAGAACTTAAGAAAGTGA 2392  
2279 AAGTAAAGAGCTTGAAGAAAGAAAGAAAGTAAAGTAAAGTAAAGAGAG 2338  
2393 AGAAGCAATGAAGCAAGTCCAGCTTGTGCTGCTCATTAAGAAAGCAAGTATG 2452  
2339 AAAAAGCATTGGAAGTGAAGTCACTTATTTATGAGAGGAAAGAGCTGATATAT- 2397  
2453 TATCAAAAGTAAATAGCAAAAGCAAGTAAAGTCCGTTCAACGGACTTCAAGTA 2512  
2398 -----AATTAATTAAGTCAATTAAGTCACTCATCTGATACCTTAAGAGAAAGGCA 2452  
2513 CCAAGAAACATGTGAAATTAAGAGAGGAGTAAAGATGATCCGTAAAGAAATTAG 2572  
2453 CAACAGAGTTT-----AACTTGAAGAAAGAAAGTCAAGTAAAGAAAG 2506  
2573 AAGCTAAGCATTTGATTTGGCAGCAGAGTATTTGGAAGATATGATTTGAAGAA 2632  
2507 AATTAGCAGCATTTGATTTGGTACCAAGAGATTTGATCTATCTAATTTGAAGAA 2566  
2633 GATGTAATTAATTTGAAATCAGATTTGAGAAATTAAGAGATTTGCAAGACTTAAGAA 2632  
2567 TATGTAATTAATTTGAAATTTGAGAAATTTGAGAAATTTGCAAGAGTCT 2623  
2693 TATGCAAAAGATTAATTAAGCTTTGCGCATCTGAAGCTCTGAGAGTGAAGCCGAC 2752  
2624 CATGTAAGAAAGTACAGGAAATATTTCAACATTAAGCCACTAAAGTGAAGCAGC 2663  
2753 AAACAGTGAAGAAAGTACCAAGCACTAAGCAACA-----ACAACACCGTTGCG 2806  
2684 AATATGTAATTAATTAAGCACTAAGCACTAAGCAACAAGCACTAAGCACTAAG 2743  
2807 ATCCGAAGGCAAGCAAGTCAATCTTTACAGCAAGCAAGCAAGTGGTTACAGACAT 2866  
2744 ACGCAAGGCAAGCAAGTCCACTCTTTACAGCAAGCAAGTGGTTACAGACAT 2803  
2867 CGAACAACAGCAAGCTCTAATCTAATCTAATCTAATCTAATCTAATCTAATCT 2926  
2804 CGACCCATTAAGCAAGTCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2863  
2927 CAACGAGGCAAGTCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2986  
2864 CGAACAAGCTGTAAGCTTACGAAGTACGAAGGAGAGAGAGAGAGAGAGAGAG 2923  
2987 TGAACCGAGTGAAGGCTGAGATGAGTGGTGAAGTGAAGTGAAGTGAAGTGAAG 3046  
2924 TGAACCGAGTGAAGGCTGAGATGAGTGGTGAAGTGAAGTGAAGTGAAGTGAAG 2983  
3047 CAATGCTTATTTCTGTTCTGATT 3069  
2984 CAATGAGCATTTCAATTCATGATT 3006

RESULT 8  
US-10-654-416-15  
Sequence 15, Application US/10654416  
GENERAL INFORMATION:  
APPLICANT: Kovacs, et al.  
TITLE OF INVENTION: Identification of a region of the major surface  
TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii  
FILE REFERENCE: 4239-58054  
CURRENT FILING DATE: US/10/654, 416  
CURRENT APPLICATION NUMBER: 2003-09-02  
PRIOR APPLICATION NUMBER: US/09/762, 724

PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: PCT/US99/18750  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: US 60/096, 805  
PRIOR FILING DATE: 1998-08-17  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 249  
TYPE: DNA  
ORGANISM: Pneumocystis carinii sp. f. hominis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(249)  
US-10-654-416-15

Query Match 6.0%; Score 185; DB 19; Length 249;  
Best Local Similarity 83.9%; Pred. No. 1.2e-26;  
Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

2821 GAATGCAATCTTACAGACAGACAGACATGGGTTACAGACATGACAGACAGACAG 2880  
1 GAGTGCAATCTTCTGACAGACAGACAGACAGTGGGTTCAAGAGCTGACCCATATGAC 60  
2881 ACGTCTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2940  
61 ACTTCTAGCATGCTCAGACAGTCAATCAATCAATCAATCAATCAATCAATCAATCA 120  
2941 AAACCAACAGTGAAGCAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAG 3000  
121 AAGCTGAGAGTGAAGCAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAG 180  
3001 GGGGTGAAGTGAAGTGGGTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060  
181 GGGGTGAAGTGAAGTGGGTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
3061 TTCATGATT 3069  
241 TTCATGATT 249

RESULT 9  
US-10-668-749A-1  
Sequence 1, Application US/10668749A  
GENERAL INFORMATION:  
APPLICANT: Agilent Technologies  
TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis  
FILE REFERENCE: 50112-1580  
CURRENT APPLICATION NUMBER: US/10/668, 749A  
CURRENT FILING DATE: 2003-09-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 1300  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: synthetic construct  
US-10-668-749A-1

Query Match 3.0%; Score 91.4; DB 20; Length 1300;  
Best Local Similarity 42.0%; Pred. No. 8.4e-08;  
Matches 536; Conservative 0; Mismatches 741; Indels 0; Gaps 0;

78 AAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 137  
1 AAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 60  
138 ATTACTGAAGCAAACTAAATATGAACAAGTACAGAAACTTAAAGTTTGGCA 197  
61 AAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 120



Db 791 AATAAANAAAAAAGAAAAAAGATACAAANAAAAAAGAAAAA 732  
Qy 574 ACATGTAAGATATTTTAAATGAAAAATAGAGTGGCTACTTTAAACGATGTT 633  
Db 731 AAAAAAAGAAAAAAGAAAAAAGATTAAGAAAAAAGAAAAAAGTAAAAA 672  
Qy 634 TCAGCAGCAGTACGAGATTTTAAAAAAGAAACATGCTTGAATTTCTGACAAATGCTAT 693  
Db 671 AAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGTAACT 612  
Qy 694 TTTTACATTTGGAATTTGGGAGCAGATTAATTAATGTTGAAATTTGGGAGAAAA 753  
Db 611 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGTAAAAA 552  
Qy 754 TGGCAGAGACAAATA 769  
Db 551 AACCATATTAATAAAA 536

## RESULT 11

US-10-221-714A-461/c  
; Sequence 461, Application US/10221714A  
; Publication No. US20040048254A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; FILE REFERENCE: 5013.1005  
; CURRENT APPLICATION NUMBER: US/10/221, 714A  
; PRIOR FILING DATE: 2003-01-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/02955  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: DE 10013847.0  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 540  
; SEQ ID NO 461  
; LENGTH: 6292  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-221-714A-461

Query Match 2.8%; Score 87; DB 19; Length 6292;  
Best Local Similarity 44.3%; Pred. No. 1.2e-06;

Matches 487; Conservative 0; Mismatches 605; Indels 7; Gaps 3;

Qy 83 AAGATGACTAGTACGAGATGCAAAAAAATCTAAAAAATTTGCCAGAAATGA 142  
Db 4475 AAAAAAATATTAATAAAAAAATTAATAATAAAAAAACAATAATATA 4416  
Qy 143 CTGAGAGAAATCTAAATATGAGACAGTACAGAAAACTTAAAGTTTTTGCAGAAATG 202  
Db 4415 AACCAAAAAAATTTTCAAAAAAATTAATAAAAAAATTTTAAATTAATAA 4356  
Qy 203 GAAAAAGAGATACAAATGCAAAAGAACTGAAGGCAATATTGAGAAAAATGTACTCAA 262  
Db 4355 AAAAAAATTAATAAAAAATTCGTAATAAAAAAAGAAAAATTAACAAAAAATTTT 4296  
Qy 263 TCAAGAGAAATCTTAAGAGCAATTAATAAAAAAATTTGAGATTAAACGATTAAGATT 322  
Db 4295 ACAAAATATATTAATAAAAAAATTAATAAAAAAATTAATAAAAAAATTTTAA 4236

Qy 323 GCAAGAGAAATGACAAATGCTATTTTGGAGGAGATATGTTCAAAAGACTTAAAG 382  
Db 4235 TAATTAATAAAAAAATTAACAAATTAATAAAAAAATTAATAAAAAAATTAATA 4176  
Qy 383 ATGATTCGAATACCTTGAGAAATTAAGTCTATCAAAAGAAACGTATTAAGTTCCGAG 442  
Db 4175 AAAAAAATTAATAAAAAAATTAATAAAAAAATTTTATTAATAAAAAAATTAATA 4116  
Qy 443 AGTCTCTT-TAAGAGACTCTGAGGATCTTAATGATGATGATGATGATGATGATGAT 501  
Db 4115 AACTTTTAATAATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAA 4056  
Qy 502 CTTAAGAGATTTGCTGATGAGGAGGAAAGTATGATGATGATGATGATGATGATGAT 561  
Db 4055 AACGACCGAAAAAATTAATAAAAAAATTTTACTTAATAAAAAAATTAATAAAAAA 3996  
Qy 562 AACGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 620  
Db 3995 AAAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3936  
Qy 621 TAAACGATGTTTACGACGACCTAGAGAGTTTAAAGAAACATGCTTGAATTAAT 680  
Db 3935 TAAAAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAA 3876  
Qy 681 CGAACATGCTATTTTACATGGAATTTGCGAGAGAGATTAATAATGATTTGA 740  
Db 3875 ATAAAAAATTAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAA 3816  
Qy 741 ATGGAGAGAAATGCGAAGAAATTTGCTTAATGATGATGATGATGATGATGATGAT 800  
Db 3815 AAAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3756  
Qy 801 TCCAACTAGGCGAGGCTCAATAGCAGAGATTAATGAGCTGAAAGTTTAAAGAA 860  
Db 3755 AAAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3696  
Qy 861 GGTAGAGAGAGATGATGATTTTATTTGAGAAATCATCTAAGAGATGCGACGCTTTGTT 920  
Db 3695 AACGACAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATAAAAAA 3636  
Qy 921 GCGATGTTGATCCAGA-----TTCTAGTCTTAAAAAAGAGAGAGAAATGTC 975  
Db 3635 AAAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3576  
Qy 976 GAAAGAGCCCTTCAAAAAAGCTGCAAAATCTCATGAACATGAGGCTTTAGAAATTTA 1035  
Db 3575 CAACAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3516  
Qy 1036 TGTAAAGAAATGATGATTAATGATGAGAAAGAAAAATGTAAGATGCAAAATGAT 1095  
Db 3515 AAAAAAATTAATAAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3456  
Qy 1096 ATTAACAAAATTTGCAAAATTTTCACTCAAAAGTCAATTAATGCTTTTGTATCA 1155  
Db 3455 AACACAGATTAATAAAAAATTTTAAATTAATAAAAAAATTAATAAAAAAATTAATA 3396  
Qy 1156 ACAGAGAGAAATTAAGAA 1174  
Db 3395 ATATTAATAAAAACTTTACAA 3377

## RESULT 12

US-10-425-115-120013/c  
; Sequence 120013, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: Ia Rosta, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21 (53222) B

```
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 120013
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: (1) (1062)
; LOCATION: (1) (1062)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_40935C.1
US-10-425-115-120013

Query Match      2.7%; Score 84.2; DB 21; Length 1062;
Best Local Similarity 45.5%; Pred. No. 2.1e-06;
Matches 296; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

Oy 78 AAAGAGATGACTAAGTGAAGCAGAAATGCAAAAAAAAACTAAATAATTTGCCAAGA 137
Db 919 AAAAAAAAAAAAAAAAAATTAATAAGAAAAAAAAAAAAAAAAAAAAAAAAAGAAAA 860

Oy 138 ATTGCTGAAGCAAACTAAATATGAAACAGTACACAGAAACTTAAGTTTGGCA 197
Db 859 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAATTA 800

Oy 198 AGATGGAAGAGAGATACAAATGCAAGAACTGAAGCAATATTGAGAAAAATGTAC 257
Db 799 AAAATTAATAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAA 740

Oy 258 TACAATCAAGAGAACTTAAGAAAGCAATTAATAAAAAAAAAATTCAGATTACGATPA 317
Db 729 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAATA 680

Oy 318 GGATTGCAAGAGATGAACCAATGCTATTTTTGGAGGAGATATGTCAAAAGACT 377
Db 679 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 620

Oy 378 TAAAGATGATGCACTACTTTGAGAAATTAAGTCTATCAAAAGAAACGTATTAAGTTGC 437
Db 619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAAAAAA 560

Oy 438 GGAAGAGCTTTCTTTAAGACACTTCGTAGCGACTTTAATGATGATCATATGTGAAGA 497
Db 559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 500

Oy 498 AAAAATTAAAGATTTGCCCTGTCTATGGGAGGAGAAAGTGTAGTTAACAACTTTGTG 557
Db 499 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAGTAAGAAATA 440

Oy 558 TCTGAACGAGAAAGACATGTAAAGATATTTTAATTGAAAAAGATAAGAGTCCGTAC 617
Db 439 AAAGATCAAGAAATTAATAAAAAAAAAAAAAAAAAATTAAGATCAAAATTAAGTAAGTAAGCC 380

Oy 618 TCTTAAACAGATGTTTACGACCACTAGGAAGTTTAAAAAAGAAACATGCTTTGAATT 677
Db 319 GAACACACATTAATTAAGACAAATAAGCAACAAATAAGACGCGCATAGTAATG 320

Oy 678 ACTGAAACATGCTATTTTACATTGGAATTCGCGAGACGACGATATA 727
Db 319 ATAGAAGAAATTTTAAAGAAACTCTACTTAACATAGATTTAAGACGA 270

RESULT 13
US-10-425-115-81106/c
; Sequence 81106, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
```

```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 81106
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (1004)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_173986C.1
US-10-425-115-81106

Query Match      2.7%; Score 82.2; DB 21; Length 1004;
Best Local Similarity 46.4%; Pred. No. 5e-06;
Matches 267; Conservative 0; Mismatches 309; Indels 0; Gaps 0;

Oy 112 AAAAACTAAATAATTTGCAAGAAATTGACTGAGCAAACTAAATATGAACAAGTA 171
Db 1004 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 945

Oy 172 CACAGAAACTTAAGGTTTTCGAAAGATGAAAGCAGATACAAATGCAAGAACTG 231
Db 944 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAA 885

Oy 232 AAAGCCATATTGGAATAAATGTACTACATCAAGAGAACTTAAGAGCAATTA 291
Db 884 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAACCAAAAAAAAA 825

Oy 292 AAAAAATTCAGATTATTAACGATTAAGATTGCAAGAGATGAACAACATGCTATT 351
Db 824 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 765

Oy 352 TTGAGGAGATGTCTCAAAAGAACTTAAGATGATGCAATCTTGAGAAATTAAGTC 411
Db 764 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAAATA 705

Oy 412 TATCAAAAGAAAGTGAATTAAGTTCGGAAGAGTTCTTTTAAGACCTTCGTAGCAT 471
Db 704 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 645

Oy 472 CTTAATGATCATATGTAATGAAAAAACTTAAGAGATTTTGCCCTGTATGCGAGAG 531
Db 644 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 585

Oy 532 GAAAGTGAAGTTTAACAACTGTGTCTGAACGAGAAAGACATGTAAAGATATTTTA 591
Db 584 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 525

Oy 592 ATTGAAAAAGTAAGAGTGGGTACTTTAAACAGATGTTTACGACGACTAGGAAGT 651
Db 524 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAATTAATAAACACAAAAAGAAAA 465

Oy 652 TTTAAAAAAGAAACATGCTTAATTAAGTAATCTGACAA 687
Db 464 CAATAAGAACAAATTAATTAATTAAGTAACAAAA 429

RESULT 14
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```

; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710
```

Query Match 2.7%; Score 81.8; DB 21; Length 921;

Best Local Similarity 45.8%; Pred. No. 5.8e-06; Matches 319; Conservative 0; Mismatches 377; Indels 1; Gaps 1;

```

QY 71 TAAATCTAAAGAAAGATGAGTAACTAGTACGAGAAATGCAAAAAAATAAATAATTT 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 846 TAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAT 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 GCCAAGATTGACTGAGCAAAATCTAAATATAGAAACAAGTACAGAAAATTAAGGTT 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 786 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 191 TTTCGGAATGGAAGCAAGATCAAAATGCAAGAACTGAAGCCATATTGAGAAA 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 726 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 251 AATGTACTACATCAAGAAATCTTAAGAGCAATTAATAAATAAATTGAGTTTAA 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 CGGATAGATTTGGAAGAGATGAACAACATGCTATTGAGGAGTATGTTCAA 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 371 AAGAACTTAAGATGATTCATCTTGAAGAAATTAAGTCTATCAAAAGAACGTGATA 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 431 AAGTTGGGAAGATTCTTTAAAGACACTTGGAGCATTTAATGATCACTCATAT 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 ATCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGT 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 GTGAAAAAATCTTAAGAGATTTGCCCTGTCATGGGAGGAAAGTGAATTAACA 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 AGAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 551 ACTGTGCTGACACGAAAGAGACATGTAAAGATTTTAAATTGAAGAAAGATTAAGT 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 AATCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 GCGTACTCTTAAACAGATGTTTCAGACGACTAGAGAGTTTAAAGAAACATGTC 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 AGAAAAAATTAACAAATTTATTCGTAAAAAATAAATAAATAAATAAATAAATAA 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 TTGAATTAATCGAATGCTAATTTTACATTTGAATTTGGGAGAGAGATTAATTA 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 731 AATGATTTGAATTTGAGAGAAATGCCAAAGAACAAA 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 AACAAAAATGTAGAAAAAATGTACAAAAATTAATA 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15  
US-10-425-115-172717/c  
; Sequence 172717, Application US/10425115

```

; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 172717
; LENGTH: 1243
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1243)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717
```

Query Match 2.6%; Score 81.4; DB 21; Length 1243;

Best Local Similarity 44.6%; Pred. No. 7.9e-06; Matches 292; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```

QY 78 AAGGAAGATGACTTAATAGTACGAGAAATGCAAAAAAATAAATAATTTGCCAAGA 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1134 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 138 ATTGACTGAAGCAAAATCTAATATATAGAAACAAGTACAGAAAATTAAGGTTTTCGA 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1074 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 198 AGATGGAAGACGATCAAAATGCAAGAACTGAAGCCATTTTGAAGAAAAATGTAC 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1014 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 258 TACATCAAGAAATCTTAAGAGCAATTAATAAATAAATTGAGTTTAAACGATTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 954 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 GATTCGAAGAGATGAACAACATGCTATTGAGGAGTATGTTCAAAAGAACT 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 894 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 TAAAGATGATTCGCAATCTTGAAGAAATTAAGTCTATCAAAAGAACGTGAATTTGC 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 834 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 GGAAGAGTTCTTTAAAGCACTTGTAGCGATCTTAATGATCACTATGTGAAA 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 774 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 498 AAAAATTAAGAGATTTGCCCTGTCATGGGAGGAAAGTGAATTAACAATCTGTG 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 558 TGTGAACGAGAAAGACATGTAAAGATTTTAAATTGAAGAAAGATTAAGAGTCCGTAC 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 618 TCTTAAACAGATGTTTCAGACGACTAGAGAGTTTAAAGAAACATGCTTGAATT 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 AGGAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 678 ACTGGAATGCTAATTTTACATTTGAATTTGGGAGAGAGATTAATTAATA 732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 ATAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: October 29, 2005, 09:48:37  
Job time : 2347 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: October 29, 2005, 06:20:41 / Search time 10576 Seconds

(without alignments)  
4686.998 Million cell updates/sec

Title: US-10-654-416-14

Perfect score: 5411  
Sequence: 1 MARAYVRQAAAGTQNSIDEH.....MSGNMVRGVIVMVISFMI 1023

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4708233 segs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US10654416/runat\_26102005\_101022\_21731/app\_query.fasta\_1.1159  
-DB=GenEmbl -QPM=faetap -SUFPIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPM=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10654416.QCGN\_1\_17019.Qrunat\_26102005\_101022\_21731 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_on:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5411	100.0	3072	6	AR438440 Sequence
2	5406	99.9	3071	8	AF033211 Pneumocys
3	4398.5	81.3	3054	6	AR438439 Sequence
4	4398.5	81.3	3056	8	AF033210 Pneumocys

5	4352	80.4	3084	6	AR438437 Sequence
6	4347	80.3	3083	8	AF033209 Pneumocys
7	3827	70.7	3090	6	AR438436 Sequence
8	3822	70.6	3089	8	AF033208 Pneumocys
9	3631.5	67.1	3363	8	PMCSGI
10	3401.5	62.9	3081	6	AR438438 Sequence
11	3396.5	62.8	3080	6	AF033212 Pneumocys
12	2762.5	51.1	3042	6	AR438434 Sequence
13	2762.5	51.1	12792	8	AF038556 Pneumocys
14	2404.5	44.4	3006	6	AR438435 Sequence
15	1731.5	32.0	3563	8	PMCSGIMS
16	1707	31.5	3511	8	PMCSGI
17	1707	31.5	36317	8	PMCSGI
18	1701	31.4	37288	8	PMCSGI
19	1687.5	31.2	11363	8	PMCSGI
20	1668	30.8	33547	8	PMCSGI
21	1658.5	30.7	11370	8	PMCSGI
22	1648	30.5	25824	8	PMCSGI
23	1648	30.5	25824	8	PMCSGI
24	1648	30.5	32542	8	PMCSGI
25	1648	30.5	35756	8	PMCSGI
26	1644	30.4	5701	8	PMCSGI
27	1628.5	30.1	3242	8	PMCSGI
28	1627.5	30.1	3016	8	PMCSGI
29	1620	29.9	3520	8	PMCSGI
30	1607	29.7	21458	8	PMCSGI
31	1586	29.3	4002	8	PMCSGI
32	1582.5	29.2	3321	8	PMCSGI
33	1581	29.2	3491	8	PMCSGI
34	1531.5	28.3	2982	8	PMCSGI
35	1502.5	27.8	3419	8	PMCSGI
36	1459	27.0	183557	2	AL590130 Homo sapi
37	1439.5	26.6	3287	8	AF063235
38	1370.5	25.3	3125	8	AF063238
39	1349.5	24.9	183557	2	AL590130 Homo sapi
40	1324.5	24.5	3459	8	AF149015
41	1302	24.1	3484	8	AF063234
42	1155.5	21.4	2180	8	PMCSGI
43	1076	19.5	2528	8	AF063237
44	1055	19.5	2370	8	PMCSGI
45	1052	19.4	2186	8	PMCSGI

#### ALIGNMENTS

RESULT 1  
AR438440  
LOCUS AR438440  
DEFINITION Sequence 13 from patent US 6664053.  
ACCESSION AR438440  
VERSION AR438440.1 GI:42663295  
KEYWORDS  
ORGANISM Unknown.  
SOURCE Unknown.  
REFERENCE 1 (bases 1 to 3072)  
AUTHORS Kovacs J.A., Huang S., Mauer H., Fischer S.H., Gill V.J. and Wei Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG) gene of human *Pneumocystis carinii*  
Patent: US 6664053 A 13 16-DEC-2003;  
JOURNAL  
FEATURES  
source 1. 3072  
1. 3072  
/organism="unknown"  
/mol\_type="genomic DNA"

#### ORIGIN

Alignment Scores:  
Pred. No.: 0  
Score: 5411.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
Length: 3072  
Matches: 1023  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-10-654-416-14 (1-1023) x AR438440 (1-3072)

QY 1 MetAlaArgAlaValIysArgGlnAlaIaGlyThrGlnAasnSerIleAspGluGlnHis 20

Db 1 ATGGCGCGGGCGGTCAAGCGGCAAGGACACAGAAATGCAATGGATGGAGAACAT 60

QY 21 ValIeuAlaLeuIleLeuIleGluAspGlyLeuSerGluGlnGluCysLeuIleLeu 40

Db 61 GTTTAGCTTAAATCTTAAGAAAGATGCACTAAGTGACAGCAATGCAAAAAAATACTA 120

QY 41 LysIleSerYrCysGlnGluLeuThrGlnAlaIleLysLeuAenIleGluGlnValHisArgIys 60

Db 121 AAAAAATTTGCAAGAAATGACGAAACCAAACTAAATATAGAACAAATGACACGAAAA 180

QY 61 LeuIysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuIleAsn 80

Db 181 CTTAAAGCTTTTGGAGATGGAAAGCAGATTCAAAATGCCAAAGAACTGAAGCCCAAT 240

QY 81 IleGluIleLysCysThrThrIleLysGlyLysLeuIleGluAlaIleLysLeuIle 100

Db 241 ATTGAGAAAAAATGATCAATCAATCAAAAGAACTTAAAGAAAGCAATTTAAAAAATTT 300

QY 101 GlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGlnGly 120

Db 301 CAGATTATACGAGATTAAGATTGCAAAAGAAATGAACAAATGCTATTTTGGAGGGA 360

QY 121 ValCysSerLysGluLeuIleLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys 140

Db 361 GTATGTTCAAAAGAACTTAAAGATGATTCGAATCTTTGAGAAATTAATGCTATCAAAAAG 420

QY 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160

Db 421 AAACCTGATTAAGTTGCCGGAAGAGTTCTTTAAAGACACTTCGTAAGCACTTAAATGA 480

QY 161 SerValIleCysGluIleLysLeuIleGlyLysCysProValMetGlyArgGluSerAsp 180

Db 481 TCAAGCATATGTAAGAAAAAACTTAAAGAAATTTGCTGTCATGGGAGGAAAGTAT 540

QY 181 GluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLys 200

Db 541 GAGTTTAAACAACTTGTCTGTAACCAAGAAAGACATTAAGAAATTTTAAATGAAAAA 600

QY 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaLeuGlySerPheLysLys 220

Db 601 GATAAGAAAGTCCGGTACTTTAAAAACAGATGTTTCAGCAGCACTGAGAAAGTTTAAAAA 660

QY 221 GluThrCysLeuGluLeuLeuGluGlnCysTyrPheTyrIleGlyAsnCysGlyAspAsp 240

Db 661 GAAACATGTCCTTAATTAAGCAATGCTATTTTTCATTTGGAATTTGCGAGACGAC 720

QY 241 AspIleIleLysCysIleGluLeuGlyLysCysGlnGluGlnAsnIleAlaIleTyrMet 260

Db 721 GATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780

QY 261 ProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGly 280

Db 781 CCACCAAGACCCGATTTTGATCCAACTAGCCAGAGGCTACCAATGACAGAGATTAAGG 840

QY 281 LeuGluGluPheTyrLysLysValIleGluGluAspGlyValPheIleGlyLysAsnHisLeu 300

Db 841 CTGGAAAGCTTTTAAGAAAGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATATTA 900

QY 301 ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysAsp 320

Db 901 AGAGATGGACAGACTTTTGGCATTTGTGATCCAAAGATTTCAAGCTTAAAAAAAAGAC 960

QY 321 AspLysGluLysCysGluGluAlaLeuGlnLysSerCysLysAsnProHisGluHisGlu 340

Db 961 GACAAAGAGAAATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

QY 341 AlaLeuGluSerLeuCysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGlu 360

Db 1021 GCTTAAAGAGTTTATGTAAAGAAAAATGCTTTAAGTAAATGATGAGCAAGAAAAATGCA 1080

QY 361 GluLeuGlnAsnAspIleAsnLysThrCysLysIlePheThrSerLysValThrAsnAsn 380

Db 1081 GAATTTGCAAAATGATTAATTAACAAACTTGCAAAATTTTCACTTCAAAAGTCACTAAATAT 1140

QY 381 ArgLeuPheAspProThrLysGlyAsnAsnGluIleValGlyTyrPglGluIleuProThr 400

Db 1141 CGTCTTTTGTATCCAAAGAAAGAAATTAATGAATTTGTGATGGAGAGGTTCCCAACA 1200

QY 401 PheLeuSerAsnGluAspCysAlaLysLeuGluGluSerYrCysPheTyrPheGluLysLys 420

Db 1201 TTTCTTGCAACGAAAGTTGTGCAAAATTTGAGAGTCCATTTGTTCTTATTTTGAAAAAA 1260

QY 421 CysProAspGlyGluAsnAlaCysLysAsnIleArgAlaThrCysTyrLysArgIleu 440

Db 1261 TGTCAGATGAGAAATATGCAATGTAATAATTAAGCAACATGTTCAAAAGAGAGACTT 1320

QY 441 AspAlaArgAlaAsnLysValLeuGlnGluAsnMetArgGlyMetLeuHisGlySerAsn 460

Db 1321 GATGACAGGCGAAATTAAGTGTGCAAGAAATTTGCGAGAAATGTTACATGTTCAAAAC 1380

QY 461 LysSerThrLeuGluLysPheGlnGlnGluLeuValLysValCysGluLysLeuLysLys 480

Db 1381 AAAAGCTGCTTGAAGAGTTTCAACAAAGATTAATTAAGATGATGAGAACTGAAAAA 1440

QY 481 GluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLys 500

Db 1441 GAAACCAAGAAAGTTTCTCAACGATGAATTAATTAATTTATGTTGTTGTAACAGCCAGAAA 1500

QY 501 AlaAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeu 520

Db 1501 GCACCCGCGTGTCAACATGATCTTCGAATGAATTAATTAATTTTAAACGCAACAACTG 1560

QY 521 AspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGln 540

Db 1561 GATCAAAAGCAAGATTTCCGCAAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCAA 1620

QY 541 AspLeuGlyGluAspSerLysGluIleThrTyrProCysHisThrLeuGluGlnGlnCys 560

Db 1621 GATTTAGAGAGATTCAAAGAAATTAATGAGCAATGTCATACATGAGACACCAATGTC 1680

QY 561 AsnArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThr 580

Db 1681 AATGCTTGGGAGTCAAGAAATTTTAAAGCAGGTTTATGATGAGAACAAAGATACT 1740

QY 581 LeuLysAspGlnGluSerCysValLysTyrLeuLysGluLysCysAsnLysTyrSerArg 600

Db 1741 TTGAAAGCCAAAGAAAGTTGTGTAATACTTAATAAGAAAGTGAATTAATGCTTGA 1800

QY 601 ArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMet 620

Db 1801 AGAGAGATGACGCTTCTCTTTGTATGTCTTCCAAACGCTACGATGAGAGCTGATG 1860

QY 621 ValLysAspValLysAspArgCysGluValPheLysLysAsnIleLysAlaSerTyrIle 640

Db 1861 GTAAAGACGTAAGACAGGTTGAAGTATTAACAAAAAATTAATAAGCTTCAATATATT 1920

QY 641 IleGluPheLeuGluAsnAsnThrAsnLysIleThrThrLeuGluArgAsnCysProSer 660

Db 1921 ATTGAATTTCTTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980

QY 661 TrpHisThrTyrCysAsnArgPheSerProAsnCysProGlyLeuThrLysGluAsnSer 680

Db 1981 TGGCATATCGATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040

QY 681 CysThrLysIleLysLysHisArgGluProPheTyrLysArgLysAlaLeuGluAspAla 700

Db 2041 TGTAACAAATCAAGAAAGCATGAGAGCCGTTCTATTAATAAGAAAGGCTTGAAGATGCT 2100

QY 701 LeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCysGluProAlaLeuLys 720

Db 2101 CTCAAGTAGAGCTTCAAGAAATTTGACTGAATTAATTAATTAATTAATTAATTAATTAAT 2160

[illegible]

SOURCE	Pneumocystis jirovecii
ORGANISM	Pneumocystis jirovecii Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
REFERENCE	Pneumocystidaceae; Pneumocystis.
AUTHORS	1 (bases 1 to 3071)
TITLE	Garbe,T.R. and Stringer,J.R. Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii
JOURNAL	Infect. Immun. 62 (8), 3092-3101 (1994)
MEDLINE	94314421
PUBMED	7518806
AUTHORS	2 (bases 1 to 3071) Mei,Q., Turner,R.E., Social,V., Kilmington,D., Angus,C.W. and Kovacs,J.A. Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region
JOURNAL	Infect. Immun. 66 (9), 4268-4273 (1998)
MEDLINE	98380374
PUBMED	9712777
REFERENCE	3 (bases 1 to 3071) Mei,Q., Turner,R., Social,V., Kilmington,D., Angus,C.W. and Kovacs,J.A. Direct Submision
AUTHORS	Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA
TITLE	Location/Qualifiers
JOURNAL	1..3071
FEATURES	/organism="Pneumocystis jirovecii" /mol_type="genomic DNA" /specific_host="Homo sapiens" /sub_species="hominids" /db_xref="taxon:42068" /clone="HumSG35" /note="derived from HIV-infected human with P. carinii pneumonia"
SOURCE	<1..3071 /gene="MSG" <1..3071 /gene="MSG" /note="surface antigen" /codon_start=3 /product="major surface glycoprotein" /protein_id="AAC34974.1" /db_xref="GI:3560519"
gene	translation="ARAVKROAAGTQNSIDEHVLAIIKEDGSEOECKKLKYCO ELEAKINIEOVHRKAFPCEDQADTKCELRANIETCTTIGKLEKAIKKKIIOIIL TDKDCKENEOCPLFBEVCSEKELDKDCNTLRNICYOKRKAIVETLRLSRDSS VICEKLKEITCPWVGRESDELTNLCNQKCTKNILEKRCGTGLTVDYSAIGSPK KETCLLEBOCYFYIGNCGDDDIKCIELDGKOEONIAAMPDPDPTREPIATVE DIGLEBFYKVEBEGVFIRGNHLRDATALALLIOSLSLKRRDKKCEBALOXSCHN PHHEBAVESLCXKNGSLSDGTCKCEBLONINTQCTFTSKYNRNLPDPTKGNNELIV GMGJLPTPLSNPDCAKLESCEFPEKKCGPENGAACNRIRATCYRGIDABARALJOEM MRGMLHSNKSMLERFOQLVKYCCEKLKENKSSFSDLEFLICVOPRAKARLLTHD RMKTIFLRQQLDRDPDTDNCKCELGRKCODIGESKEITWPCHTLDEQCNRLTGE ILKVOILDEHKHTLDKOESCVXYLKEKCNMSRGDRFSFVCFONATCELVWDVVK DLREVFKNIKASYIIIELENNTNKITLERNCPSMTTYCNRFSPNGPLTKENSCTRK IKKHREFYKRKALEDAIKVELQGKLDKSCBPALKRYCTVAGVNNAISIGLKRAAE TKNSGKSDPAKKEICEKVKEBECRALPEIQPADIKKDYTELKRRARE AMNKSIVLSLTIKKNBSNVKSNKNDKRAVNSGLODTPTRHYKILRGKYDSYTEL EEAFVETLAAEVFGRIYVDLKERCNLESDCKIKDDCKDLSEVCKINKACNKLPLEVK PHEAVESTTTTITTTTTTVDAPKATEKCSQKIDHDWTWTOSTHSTSIISTITSKIT LSTRACKPTKCTTGEBDDAGDVYPSGELMSGMVNRGVIVAMVISFWI"
CDS	
ORIGIN	
Alignment Scores:	
Pred. No.:	0
Score:	5406.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	99.91%
DB:	8
Gaps:	0
Length:	3071
Matches:	1022
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

```
US-10-654-416-14 (1-1023) x AF033211 (1-3071)
QY 2 A1aargalavalylsarglnaleaglthrglnaenserlleaspgluglunhival 21
Db 3 GCGGGGGCGGTCAAGCGGCGAGGACAGCAAGATAGCATTTGAGAGGACATTT 62
QY 22 LeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLysLysLys 41
Db 63 TTAGCTTAAATTTCTTAAGAGAGATGCACTTAAGTGCAGAGATGCAGAAAAAACTTAAA 122
QY 42 LysTrpCysGlnGluLeuThrGluAlaLysLeuAsnIleGluGlnValHisArgLysLeu 61
Db 123 AAATATGTCAGAAATTAAGTGCAGAACAACTAAATATAGAACAGATGCAGAAAACTT 182
QY 62 LysGluPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsnIle 81
Db 183 AAAGCTTTTTCGAGAGATGCAGAAAGCAATACAAAATCCAAAAGACTAAATTT 242
QY 82 GluLysLysCysTrpThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIleGln 101
Db 243 GAGAAAAAATGTACTACATCAATCAAGAAAACTTAAAGAACATTTAAAAAAATTTGAG 302
QY 102 IleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGluGlyVal 121
Db 303 ATTATTAACGAGTAAAGGATTCAGAAAGAAATGAAACAATGCTATTTTTCGAGGAGTA 362
QY 122 CysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTrpGlnLysLys 141
Db 363 TGTTCAAAAGAACTTAAAGATGATTCGCAATCTTTGAGAAATTAAGTGTATCAAAAAGAA 422
QY 142 ArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLysAsnLysSer 161
Db 423 CGTGATAAAGGTGCGAGAGAACTTTTAAAGAGACCTTCGAGCACTTAAATGAGATCA 482
QY 162 ValIleCysGluLysLysLysLysGluIleCysProValMetGlyArgGluSerAspGlu 181
Db 483 GTCATATGTGAAAAAACTTAAAGAGATTTGCGCTGTCATGGGAGGAAAGTATAG 542
QY 182 LeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLysAsp 201
Db 543 TTTAACAACTTGTGTCTGACCAAGAAAGACATGTAAGATATTTTAATGAAAAAGAT 602
QY 202 LysLysCysGlyThrLeuLysThrAspValSerAlaAlaLeuGlySerPheLysLysGlu 221
Db 603 AAGAAAGTCGGTACTTTAAACAGATGTTTCAGCAGCACTAGAGAGTTTAAAAAAA 662
QY 222 ThrCysLeuGluLeuLeuGluGlnCysTrpPheTrpIleGlyAsnCysGlyAspAspAsp 241
Db 663 ACATGCTTGAATTAAGTCAACATGCTATTTTACATTTGAAATTCGAGAGACAGAT 722
QY 242 IleIleLysCysIleGluLeuGlyGlyLysCysGlnGluAsnIleAlaTrpSerPro 261
Db 723 AATAATTAATGATTAATGATGGAGAGAAATGCCAAAGAACAAATTAATGCTTAATGCA 782
QY 262 ProGlyProAspPheAspProThrArgProGluAlaThrIleAlaIleAspTrpIleLys 281
Db 783 CCAGAGCCCGGATTTTGATCCAACTAGCCAGAGCTTACATAGCAGAGATATAGGCTG 842
QY 282 GluLysPheTrpLysLysValGluGluAspGlyValPheIleGlyLysAsnIleLeuArg 301
Db 843 GAAAGCTTTATTAAGAGGTAAGAGAGATGAGTTTATTTAGAAACATCATCTAAGA 902
QY 302 AspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAsp 321
Db 903 GATGGAAGAGCTTTGCTTGGCATTTGATCCAAAGTCTTAAGCTTAAAAAAAAGACAGAC 962
QY 322 LysGluLysCysGluGluAlaLeuGlnLysSerCysLysAsnProHisGluAla 341
Db 963 AAAGAGAAATGAGAAAGAGCCCTTCAAAAAGCTCAAAAATCCATGAACATAGAGCT 1022
QY 342 LeuGluSerLeuCysLysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGluGlu 361
Db 1023 TTAGAAAATTATCTAAGAAAAATGCTTTAAGTATGATGGAAGAAAAATGTAAGAA 1082
QY 362 LeuGlnAsnAspIleAsnLysThrCysLysIlePheThrSerLysValThrAsnAsnArg 381
Db 1083 TTGCAAAATGATATTATTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCACTAATATGCT 1142
QY 382 LeuPheAspProThrLysGlyAsnAsnGluIleValGlyTrpGluGluLysProThrPhe 401
Db 1143 CTTTATTAATCCMAAGAAAGAAATTAATGAATTTTGGATGGGAAGGTTCCAAACATTT 1202
QY 402 LeuSerAsnGluAspCysAlaLysLeuGluLysTrpCysPheTrpPheGluLysLysCys 421
Db 1203 CTTAGCAACAGAGATTTGTCGAATTTGAGTGCCTATTTGCTTATTTTGAABAAAAAGT 1262
QY 422 ProAspGlyGluAsnAlaCysLysAsnIleArgAlaThrCysTrpLysArgGlyLeuAsp 441
Db 1263 CCAGATGAGAGAAATGATGTAAATTAATTAAGACACAATGTTCAAAAAGAGACTGAT 1322
QY 442 A1aargalaaenlyvalleuGlnGluAsnMetLysGlyMetLeuHisGlySerAsnLys 461
Db 1323 GCAGGGCAATAAAGTGTGCAGAAAAATATGCAAGAAATGTATCATGTTCAAAACAA 1382
QY 462 SerTrpLeuGluLysPheGlnGlnGluValLysValCysGluLysLeuLysLysGlu 481
Db 1383 AGCTGCTTGAAGAACTTCAACAGAAATTAAGTAAAGTATGAGAAATGAAAGAA 1442
QY 482 AsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAla 501
Db 1443 AACAAAGAGATTTCTCAACAGATGAATTTATTTATCTGTGTGTATACAGCCAGAAAAACA 1502
QY 502 A1aargleuLeuThrHisAspLysArgMetLysThrIlePheLeuArgGlnGlnLeuAsp 521
Db 1503 GCCCGTTGCTTACACATGATCTTCGATGAAGAACTATCTTTTAAACGACAACTGAT 1562
QY 522 GlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGlnAsp 541
Db 1563 CAANAGCAATTTCCGACAGATMAAAATTTGCAAGAAATTTGGGAGAAATGCCAAGAT 1622
QY 542 LeuGlyGluAspSerLysGluIleThrTrpProCysHisThrLeuGlnGlnCysAsn 561
Db 1623 TTAGAGAGGATTCAAAAGAAATTAACATGGCCATGTCTATACATGAGACAAATGCAAT 1682
QY 562 ArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeu 581
Db 1683 CGCTTGGGAGTACAGAAATTTTAAACAGAGTTTATTTAGATGAACAAAGATPACTTGT 1742
QY 582 LysAspGlnGluSerCysValLysTrpLeuLysGluLysCysAsnLysTrpSerArgArg 601
Db 1743 AAAGACCAAGAAATTTGTGTAAATTAACCTAAABAAAAGCTTAATTAATGCTTAAGAA 1802
QY 602 GlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetVal 621
Db 1803 GGAGATGACCGTTCTCTTTTGTATGTGTTCCAAAACGCTAGCGTGAAGTATGCTA 1862
QY 622 LysAspValLysAspArgCysGluValPheLysLysAsnIleLysAlaSerTrpIleLys 641
Db 1863 AAAGACGTGAAGACAGGTGTGAAGTATTCAAAAAAATATTAAGCTTCAATAATTAATT 1922
QY 642 GluPheLeuGluAsnAsnThrAsnLysIleThrThrLeuGluArgAsnCysProSerTrp 661
Db 1923 GAATTTCTTGAATTAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1982
QY 662 HisThrTrpCysAsnArgPheSerProAsnCysArgProGlyLeuThrLysGluAsnSerCys 681
Db 1983 CATAGATATTCATTAATGATTTTTCATCTAATTTGTCAGAGTCTTAAGAAAGATAGTGT 2042
QY 682 ThrLysIleLysLysHisArgGluProPheTrpLysArgLysAlaLeuGluLysAspAlaLeu 701
Db 2043 ACAAAAATCAAGAAAGCATCGAGCCGTTCTATTAAGAAAGAGCCCTTGAAGATGCTCTC 2102
QY 702 LysValGluLeuGlnGlyLysLeuThrAspLysSerLysCysGluProAlaLeuLysArg 721
Db 2103 AAAGTAAAGCTTCAAGAAATTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2162
```

Qy 722 TycTerThrVala1aagiYaenValaAsnAAsnAlaSerIleSerGlyLeuCyblyAsn1aAsn 741  
Db 2163 TATTGTACAGTAGCGGAAACGTTAAATATGCGCAATCAGTGGCTTATGCAAGACTAAC 2222  
Qy 742 ThrlyAspAsnSerGlyLySerSerPduAsp1aAarglybgiLeuCybgiLybLeu 761  
Db 2223 ACCAAGATTAACCTCGAAGAAAGATGAGGAGTCTGAAAGAAAGAACTCTGAGAAATTA 2282  
Qy 762 VallybgiVala1aagiLuglInCyblybVala1aLeuProThrgluLeuGlyInProAla1a 781  
Db 2283 GTGAAGAAAGTGAAGAAAGAGTGCAGCAATTAACAAGATTAGACAACCGCGCAGCT 2342  
Qy 782 AspLeuLyblybAspPylLybThrThrgluLuglInLeuLyblybVala1aMet 801  
Db 2343 GATCTAATAAAGATTTATAGACATATAGAGAACTTAAAGAACTGCAAGAGAAAGCAATG 2402  
Qy 802 AsnLybSerSerLeuVala1LeuSerLeuIleLyblybAsnGluSerAsnVala1SerLybSer 821  
Db 2403 AACAAAGTCCAGTCTGTTTGTCTCACTATTAAAGAAAAAGAAAGTAAATGTATCAAAAAAGT 2462  
Qy 822 AsnSerLybAsnLybAspLybAsnAlaVala1SerAsnGlyLeuGluAspThrThryHis 841  
Db 2463 AATAGCAAAAAACAAGATTAAGAAATGCGCTTCAACCGACTTCAAGATACCAACAAACAT 2522  
Qy 842 VallybIleLeuAargyVala1LybAspVala1SerVala1ThrgluLeuGluAla1aVala1a 861  
Db 2523 GTGAAGAAATTAACCGAGGAGGAGTTAAGAGATGATCCGTAAAGAAATTAAGAACTTAAGCA 2582  
Qy 862 PheAspLeuVala1aagiVala1PheGlyAargyVala1AspLeuLybgiLysCybAsnLyb 881  
Db 2583 TTTGATTTGGCAGAGAAAGATTTGGAAGATATGTAAGTTTGAAGAAAGATGTAAATTA 2642  
Qy 882 LeuGluSerAspCybAarglyIleLybGluAspCyblybAspLeuGluVala1CyblybLyb 901  
Db 2643 TTGGAATCAAGATTCAGAAATTAAGAGAGATTGCAAGAGCTTAAGAAAGATGCAAAAG 2702  
Qy 902 IleAsnLybAla1aCybAsnLeuLybProLeuGluVala1LybProHsgluThrVala1Thr 921  
Db 2703 ATTATATAGGCTTGTGCAATCTGAAGCTCTGAGGAGTGAAGCCGCAAGAAAGCTGA 2762  
Qy 922 GluSerThrThrThrThrThrThrThrThrThrVala1aAspProLybAla1aThrglu 941  
Db 2763 GAAGATCAAGCAACATCAACAACAACAACCGTTCGATCCGAAGCGCAACGGA 2822  
Qy 942 CysLybSerLeuGluInThrAspThrTrpVala1ThrgluInThrSerThrHsrThrSerThr 961  
Db 2823 TGCAAATCCTTACAGACAACAGACATGGGTTACACAGACATCGACACACAACAGCAG 2882  
Qy 962 SerThrIleThrSerThrIleThrSerLybIleThrLeuThrSerThrAargCybLyb 981  
Db 2883 TCTACTATCACTTACATCACTCAAAAAATTAATTGACATTAACGAGCGGATGCAAA 2942  
Qy 982 ProThrLybCybThrThrglyGluGluAspAsp1aagiAspVala1LybProSerGluGly 1001  
Db 2943 CCAACCAAGTGTACAGACAGAGAGAGATGATCAAGAGACGTGAACCAAGAGAGG 3002  
Qy 1002 LeuAargMetSerGlyTrpAsnVala1MetAarglyVala1aMetVala1aSerPhe 1021  
Db 3003 CTGAGGATGAGTGGTGAATGTATGAGGCGGTGATAGCAATGCTTATTCCTTC 3062  
Qy 1022 MetIle 1023  
Db 3063 ATGATT 3068

RESULT 3  
LOCUS AR438439 3054 bp DNA linear PAT 20-FEB-2004  
DEFINITION Sequence 11 from patent US 6664053.  
ACCESSION AR438439  
VERSION AR438439.1 GI:42663294  
KEYWORDS  
SOURCE Unknown.

ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3054)  
AUTHORS Kovacs, J.A., Huang, S., Masur, H., Fischer, S.H., Gill, V.J. and Mei, Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG)  
JOURNAL gene of human Pneumocystis carinii  
Patent: US 6664053-A 11 16-DEC-2003;  
FEATURES  
source location/qualifiers  
1..3054  
/organism="unknown"  
/mol\_type="genomic DNA"

ALIGNMENT Scores:  
Pred. No.: 2,256-274 Length: 3054  
Score: 4398.50 Matches: 838  
Percent Similarity: 88.81% Conservative: 75  
Best Local Similarity: 81.52% Mismatches: 98  
Query Match: 81.29% Indels: 17  
DB: Gaps: 9

US-10-654-416-14 (1-1023) x AR438439 (1-3054)

Qy 2 AlaargAlaVala1ybaArglna1aagiYThrgluAsnSerIleAspGlu--His 20  
Db 1 GCGGGGCGGTCAACGGCGGAGTAACAGAGCATCAGGCAATGATGATGAAGTGAAT 60  
Qy 21 ValLeuAlaLeuIleLeuLybgiLysLeuSerGluGluInuCyblybLybLeu 40  
Db 61 ATTTTGGCGTTGATTTCTACAAAGAGATGCAATGAAATCAAAATGCAAAAAAGTTTA 120  
Qy 41 LybLybThrCybGluGluLeuThrgluAla1aLybLeuAsnIleGluGluVala1aArgLyb 60  
Db 121 GAAGAAATCTGCGAAGGTTGAAAAAGCACTACAGACATGAGAAAAAGTACATTAATG 180  
Qy 61 LeuLybGlyPheCybgiLysLeuLyb-----LybAlaAspThrLybCybLybGluLeu 77  
Db 181 CTTAAAGATTTCTGTGGAATGGAATGGAAGCAAGTAAACAAATACAAATATGCAAGTCTA 240  
Qy 78 LybAlaAsnIleGluLybLybCybThrThrIleLyb---GlyLybLeuLybGluAla1a 96  
Db 241 CAAGCCAAATTTGCGGGAATGTACAAATTTTAAACACAAAGCTTAGACCAAGCTTA 300  
Qy 97 LybLybLybIleGluIleIleThrAspLybAspCybLybGluAsnGluGluInuCybLeu 116  
Db 301 ACAAAATCCA-----TCAGATGATTAATTCGAAAGAGAGTGAACCAATGCTTA 348  
Qy 117 PheLeuGluGlyVala1CybSerLybGluLeuLybAspAspCybAsnThrLeuAargAsnLyb 136  
Db 349 TTTTGGAGGAGAGCATG---CATTAATCTGTGAAGATTTGTAAACAATTAAGAAATCTA 405  
Qy 137 CysTrgIlnuLybLybAargAspLybVala1aagiGluVala1LeuLeuAga1aLeuAargSer 156  
Db 406 TGTTAACCAAGAAAAACCTGACGAGTACGAAAGAACTCTTTTGAAGGGAATTCCTGAGT 465  
Qy 157 AspLeuAsnGlySerVala1IleCybGluLybLybLeuLybGluIleCybProVala1MetGly 176  
Db 466 GATCTCAATTAACACAGAAACATGAAAGAAAGATGAAGAGATTTCACAGTCTTCAG 525  
Qy 177 ArgGluSerAspGluLeuThrAsnLeuCybLeuAsnGluLybGluThrCybLybAsnIle 196  
Db 526 AGGGAAGTAAATGATTAACGAGACTGTGTGAAACCAAGAAAAAGAGCTGCAAGATATT 585  
Qy 197 LeuIleGluLybAspLybLybCybGlyThrLeuLybThrAspVala1SerAla1aLeuGly 216  
Db 586 ATAAAGAAAAAGATTAATAATGACATCACTTAAGCAATGTTGCAACAGCACTTGA 645  
Qy 217 SerPheLybLybGluThrCybLeuGluLeuGluGluInuCybTrpPheTrgIleGlyAsn 236  
Db 646 AGTTTAAAAAGAAATATGCTTTGAATTACTTGAACAAAGCTTAATTTTCAATTGCAAT 705  
Qy 237 CybGlyAspAspIleIleLybCybIleGluLeuGlyGlyLybCybGluGluGluAsn 256

Dh 706 TCGGAGACGACGATATTAATTAATGCTATTGGAATTCGGAGGGAATGCCAAGAACAAAC 765  
Oy 257 ILeaIaTywecProProGlyProAspPheAspProThraTyrProGluAlaThrIleAla 276  
Dh 766 ATGTTGTTATATACCAACGAGACCGCATTTTGATCCAACTAGACGAGAGCTTACACTAGCA 825  
Oy 277 GluAspIleGlyLeuGluGluPheTyrIlySlyValIguGluAspGlyValPheIleGly 296  
Dh 826 GAGGACATAGACCTGGATGAGCTTTATATAAAGGACGAGAGAGAGGATGGTTTTATTTGGA 885  
Oy 297 LysAsnHISLeuArxAspAlaThrAlaLeuAlaLeuLeuIle---GlnAspSerSer 315  
Dh 886 AAACATCTTTTAAGAGATGCGACACCTTTATGACGTTGTTGTTAAAGAAAGATATCA 945  
Oy 316 LeuIlySlyAspAspIlySglIlySyrCysGluGluAlaLeuGlnIlySerCysIlyAsn 335  
Dh 946 GGAATAAATAATTAATATATCGAGAAATAATGCAATTAATTCGCAAGATAAATGCAAAAAAC 1005  
Oy 336 ProHISgluHISgluAlaLeuGluSerLeuCysIlySlyAsnGlyLeuSerAsnAspGly 355  
Dh 1006 TCTCAACGACATGAAAGCTCTAAATAATTTATGTATTAATTAATAGCTCTTAATGCAATATGGA 1065  
Oy 356 ThrIlySlySyrGluGluLeuGlnAsnAspIleAsnIlyThrCysIlyIlePheThSer 375  
Dh 1066 AAAGAAAATGCCAAAGAAATTAGAACAGATATTAAAAAAACATGCACAAACCTCAAAACA 1125  
Oy 376 LysValIThrAsnAsnArgLeuPheAspProThraIlySglAsnAsnGluIleValIglYTrp 395  
Dh 1126 ACGATTCTTAAAAACCATCTTATGATCCA-----AATGATAAATAATGTTGATAGTGG 1176  
Oy 396 GluGlyLeuProThraPheLeuSerAsnGluAspCysAlaIlySleuGluSerTyrCysPhe 415  
Dh 1177 AGAAAACCGCAACATTTCTTAATAATGAGACTGTGCAAGATTTGGAATCTTATATGTTTT 1236  
Oy 416 TyrPheGluIlySlySyrProAspGlyGluAsnAlaCysIlyAsnIleArxAlaThrCys 435  
Dh 1237 TACTACGAAAAAGCTTGTCCAAAATGCCAAAGAGGTATGAAATTTGAGGCGCGTGT 1296  
Oy 436 TyrIlyAspGlyLeuAspAlaArgAlaAsnIlyValIleuGlnIlyAsnMetArgIlyMet 455  
Dh 1297 TATAGAGAGGGCTTATGAGACGGGCATAATAAAGTCTGCAAGAAATATGCTGGGTTA 1356  
Oy 456 LeuHISglYSerAsnIlySerTTrpleuGluIlySyrPheGlnGlnIlyLeuValIlySValCys 475  
Dh 1357 TTACCTGGTTCAAAATCAAGTTGGCTTAAGAGTTTCAACAAAGATTGTAAAGATAAGT 1416  
Oy 476 GluIlySleuIlySglIlyAsnIlySglYSerPheSerAsnAspGluLeuPheIleLeuCys 495  
Dh 1417 AAGAGCTH---AAAGAAAATTAAGAAAGTTTCCAAACGATGAATAATTTGTTCTGTGT 1473  
Oy 496 ValGlnProAlaIlyValaIaArgLeuLeuThraIISAspLeuArgMetIlyThrIlePhe 515  
Dh 1474 GTACAGCCAGCAAAAGCTGACAGATTACTTACACAGATCATCAATGACGAGTTATCTTT 1533  
Oy 516 LeuAspGlnGlnIlyAspGlnIlySyrArgAspPheProThraSlyAsnCysIlySglIlyLeu 535  
Dh 1534 TTACACACAAATTTGATCAAAAGAGATTTTCCGACAGATAAAGACTGCAAGGAATTA 1593  
Oy 536 GlYArGlySyrGluAspLeuGlyGluAspSerIlySglIlyIleThrTriProCysHISThr 555  
Dh 1594 GGGAAAAAATGCCAAGATTTAGAGAAAGATTCAAAAGAAATTAACATGCGCATGTCAATACG 1653  
Oy 556 LeuGluGlnGlnIlyAsnAspArgLeuGlyIlyThrIrgIlyIleuIlySglIlyValIleuAsp 575  
Dh 1654 CTGGAGACGACAAATGCATCGCTTGGGACTACAGAAATTTTAAACACAGTTTATTTGAT 1713  
Oy 576 GluHISlyAspThrIleuIlyAspGlnIlySerCysValIlySyrIlyLeuIlySglIlySyrS 595  
Dh 1714 GAACCAAAAGATTACTTTGAAAGACCAAGAAAGTTGTGTAATACTTAAAGAAAAAGGT 1773  
Oy 596 AsnIlySTPSeArxArgIlyAspAspArgPheSerPheValCysValPheGlnAsnAla 615  
Dh 1774 AATAAATGCTTACAGAGAGATACACGTTTCTCTTTTGTATGTCTTCCAAAACGCT 1833

Oy 616 ThrCysGluLeuMetValIlyAspValIlyAspArgCysGluValIlyPheIlySlyAsnIle 635  
Dh 1834 ACGTGTAGAGCTGATGATGAAAAAGAGTGAAGAAGCGGTGTGAATATTCAAAATAATTA 1893  
Oy 636 LysAlaSerTyrIleIleGluPheLeuGluAsnAsnThrAsnIlySylIleThrIleuGlu 655  
Dh 1894 AAAGCTTATATATTAATTAATTAATTTCTTGAAATATATCAAAATTAATAACACACTGGA 1953  
Oy 656 ArgAsnCysProSerTTrpHISThrTyrCysAsnArgPheSerProAsnCysProGlyLeu 675  
Dh 1954 AGAAATTTGCTCTTGGCANTAGATTTGGCAATAGATTTTCACTTAATTTGCCAGGTCTT 2013  
Oy 676 ThrIlySgluAsnSerCysThrIlySylIleIlySlyHISArgGluProPheTyrIlyAspGlyS 695  
Dh 2014 ACGAAAGAGATATGTTTACAAAAATCAAGAAAGATTTGTGACCGCTTCTTAATAAAGAAAG 2073  
Oy 696 AlaLeuGluAspAlaLeuIlySyrValIguLeuGlnIlySlySleuThraAspIlySerIlySyr 715  
Dh 2074 GCTTGGAAAGATGCTCTCAAAAGTAGACCTTCAGGAAATTTGACTGATTAATCTTAATGT 2133  
Oy 716 GluProAlaLeuIlySArGlyTyrCysThrValaIagIlyAsnValaAsnAlaSerIleSer 735  
Dh 2134 GAACCTGATTTGAAGAATATNTGTACAGTAGCGGGAACGTAATATAGCTCAATCATGCT 2193  
Oy 736 GlyLeuCysIlyValaAsnThrIlyAspAsnSerGlyIlySyrAspGluAspAlaArgIlyS 755  
Dh 2194 GGGTTATGCAAAAGCTTAACACCAAGATTAACCTGGAAGAAGATGTAGAGATGCTAGAAAG 2253  
Oy 756 GluIleuCysGluIlySleuValIlySglIlyValIguIlyGlnCysIlySlyValaIleuProThraGlu 775  
Dh 2254 GAACCTGTGTGAATAATTAATGTAAGAAAGTGAAGAAACAGTGCAAAGCATTTACCAACGAA 2313  
Oy 776 LeuGlyGlnProAlaAlaAspLeuIlySlyAspTyrIlySyrThrTyrGluGluLeuIlySlyS 795  
Dh 2314 TTAGAGAACCCGACGCTGATTTAAAAAAGATTAATAAGACATATGAGAACTTAAGAAA 2373  
Oy 796 ArgAlaGluGluValaMetAsnIlySerSerIleuValIleuSerLeuIlyIlySlyAsnGlu 815  
Dh 2374 CGTCAGAGGAAAGCAATGACAACTCCAGTCTTTTGTGTCACTCATTAACAAAAACCAA 2433  
Oy 816 SerAsnValISerIlySerAsnSerIlyAsnIlyAspIlySyrAsnAlaValISerAsnGlyLeu 835  
Dh 2434 AGTATGATATCAAAAGATTAATGCAAAAACAAGATTAAGAAATGCGTTTCAAGGACTT 2493  
Oy 836 GluAspThrThrIlySylIleValIlySleuArgArgIlyValIlyAspValISerValIlyThr 855  
Dh 2494 CAAGATGCCAATAAATCTGTAATAATCTACGAGAGAGATTAAGATGTATCCGTAACTA 2553  
Oy 856 GluLeuGluValaIlyValaIlePheAspLeuAlaIagluValaIlyPheGlyArgTyrValaIlePhe 875  
Dh 2554 GAATTAGAACTTAAGCAATTTGATTTGGCAGACAGAAATTTGGAAAGATATGTGATTTG 2613  
Oy 876 LysGluArgCysAsnIlySleuGluSerAspCysArgIlyIlySglIlyAspCysIlyAspLeu 895  
Dh 2614 AAGAAAGATGTAATTAATTTGGAATCAGATTGCGAATTAAGAGATTTGCAAGACTTA 2673  
Oy 896 GluGluValaCysIlySlySylIleAsnIlyAlaCysArgAsnLeuIlyProLeuGluValaIlyS 915  
Dh 2674 GAABAAGATTCAAAAAGATTAATTAAGCTTGTGCAATCTGAAAGCCCTGGAGGTGAAG 2733  
Oy 916 ProHISgluThrValIlyThrGluSerThrThrThrThrThrThrThrThrValaIla 935  
Dh 2734 CCGCACGAAACAGACGAGAAAGATCAACGCAACTACACAAACAAACAAACGCTTGC 2793  
Oy 936 AspProIlySylAThrGluCysIlySerIleuGlnIlyThrIraPheThrIlyValaIlyThrGlnThr 955  
Dh 2794 GATCGAAAGGCAACGAAATGCAAAATCTTACAGCAACAGACACATGGGCTTACACAGCA 2853  
Oy 956 SerThrIsthrSerThrSerThrIleThrSerThrIleThrSerIlyIleThrIleuThr 975  
Dh 2854 TCGACACACACACACACGTCTACATCACTACATCACTACATCAATAAATAACATTGACA 2913



Qy 976 SerThrArgArgCysGlyProCysThrCysThrArgGlyGluAaPaPaAaGlyAsp 995  
Db 2914 TCAACGAGGGGATCCAAACCAACCAAGTGTACGACAGG-----GATGATGCCAAGAAC 2967  
Qy 996 ValIleProSerGluGlyLeuArgMetSerGlyTyrPheValMetArgGlyValIleVal 1015  
Db 2968 GTGAAGCAAGTGAAGCTTGAAGGCTGACGGGTGAATGATGAGAGGGGGTATGTA 3027  
Qy 1016 AlMetValIleSerPheMetIle 1023  
Db 3028 GCAATGGTATTCTGTCATGAT 3051

RESULT 4  
AF033210 3056 bp DNA linear PLN 10-SEP-1998  
LOCUS DEFINITION Pneumocystis carinii f. sp. hominis clone HUMSG33 major surface glycoprotein (MSG) gene, partial cds.  
AF033210  
ACCESSION AF033210.1 GI:3560516  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Pneumocystis jirovecii  
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
1 (bases 1 to 3056)  
Garbe, T.R. and Stringer, J.R.  
Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii  
Infect. Immun. 62 (8), 3092-3101 (1994)  
94314421  
MEDLINE 7518806  
JOURNAL 2 (bases 1 to 3056)  
PUBMED  
AUTHORS Mei, Q., Turner, R.E., Sorial, V., K Livingston, D., Angus, C.W. and Kovacs, J.A.  
TITLE Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region  
Infect. Immun. 66 (9), 4268-4273 (1998)  
98380374  
MEDLINE 9712777  
JOURNAL 3 (bases 1 to 3056)  
PUBMED  
AUTHORS Mei, Q., Turner, R., Sorial, V., K Livingston, D., Angus, C.W. and Kovacs, J.A.  
TITLE Direct Submission  
Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA  
FEATURES  
source  
1.3056  
/location/Qualifiers  
/organism="Pneumocystis jirovecii"  
/mol\_type="genomic DNA"  
/specific\_host="Homo sapiens"  
/sub\_species="hominis"  
/db\_xref="taxon:42068"  
/clone="HUMSG33"  
/note="derived from HIV-infected human with P. carinii pneumonia"  
1.3056  
/gene="MSG"  
1.3056  
/gene="MSG"  
/notes="surface antigen"  
/codon\_start=3  
/product="major surface glycoprotein"  
/protein\_id="AAC34973.1"  
/db\_xref="GI:3560517"  
/translation="AAVAVKQVTSAGSGQYDENVILALILQEDAMEDTKCKSLKCYC  
BELKASLDMEKVKHMLKDFGNGKASKANTKCGIQAATGCTNFKTKQAGPALTN  
PSDNCESEROCFLGAGCNLYEDCNLRNLCTYQKRGVAVEVILRLRSNLT  
ETHKRLKEICPVLDRESNELTDLNQLQKTCENIYKDKKCTTLAAVATAGSK  
KEICLLEDCQFYIGNCGDDIICKIELGKCOBONIYIIPRPDPPTREPTLAE  
DIDDELTKAEBDGVFIGKHLADATALLLVKCDGTGANNIIGKCKLIEDCK  
NSQHBALKNLNNNSPNAVYGEKCELEEDIKKTCTNLKPTILKNHLYDPNDKIVEM

ALIGNMENT Scores:  
Pred. No.: 2,25e-274 Length: 3056  
Score: 4398.50 Matches: 838  
Percent Similarity: 88.81% Conservative: 75  
Best Local Similarity: 81.52% Mismatches: 98  
Query Match: 81.29% Indels: 17  
DB: 8 Gaps: 9  
US-10-654-416-14 (1-1023) x AF033210 (1-3056)  
Qy 2 AlAargAlaValIySargGlnAlaAlaGlyThrGlnAsnSerIleAapGluGlu---His 20  
Db 3 GCGGGGGGGGTCAAGCGGAGGTAAACAGAGCATCAAGGCAATATGATGATGAAGTGAAT 62  
Qy 21 ValLeuAlaLeuIleLeuIySgluAapGlyLeuSerGluGluGluGluGluGluGlu 40  
Db 63 ATTTGGCGTTGATTCTTCAAGAAAGATGCAATGAAATGCAAAATGCAAAATGTTTA 122  
Qy 41 LyeIySerTyrCysGlnGluLeuThrGluAlaLyeLeuAsnIleGluGlnValIleArgLys 60  
Db 123 GAAATATCTGCGAAGGTTGAAAGAAAGCATCTAGTAACTGAAAGTAACTAAATATG 182  
Qy 61 LeuIySglIyPheCysGluAapGly-----LysIleAspThrIySglIySglIyLeu 77  
Db 183 CTTAAAGATTCTCTGTGAAATGGAAATGGAAGCAAGTAAACAAATGCAAAATGCAAGTCTA 242  
Qy 78 LyeAlaAsnIleGluIySglIySglIyThrIleLys---GlyIySglIySglIyAlaIle 96  
Db 243 CAAGCCAAATTTCCGGGAAATGTACAAATTTTAAACACAAAGCTAGGACCGCTTA 302  
Qy 97 LyeIySglIySglIleGlnIleIleThrAapIySapCysIySglIyAaGlnGlnGlnCysLeu 116  
Db 303 ACAATGCA-----TCAGATGATATGTCGAAAGAGCTGAAACCAATGCTTA 350  
Qy 117 PheLeuGluGlyValCysSerIySglIyLeuIySapAapCysAaThrIleuArgAsnLys 136  
Db 351 TTTTGGAGGAGGAGATGCG---CATATCTTGTGAAGATTTGTAACTAAGAAATCTA 407  
Qy 137 CysTyrGluIySglIySapIySglIySglIyValIleGluGluValIleLeuAlaGlnIleuSer 156  
Db 408 TGTTCACGAGAAACCTGACGAGGTGACGAGAAAGATCTTTTGAAGGACCTTCGATG 467  
Qy 157 AspLeuAsnGlySerValIleCysGluIySglIySglIyLeuIySglIyCysProValMetGly 176  
Db 468 GATCTCAATAAACAGAAACACATGAAAGAAAGTGAAGATTTCCCAAGTCTTGACG 527  
Qy 177 ArgGluSerAapGluLeuThrAsnIleuCysLeuAsnGlnIySglIyGluThrCysIySapIle 196  
Db 528 AGGGAAGTAAATGAATTAACGAGACTTGTGTTGAACAGAAAGACGTCGAGAAATATT 587  
Qy 197 LeuIleGluIySapIySglIySglIyThrLeuIySglIySglIySerAlaIleGluGly 216  
Db 588 ATAAAGAAAAAGTAAAGTAAATGCACTCTTAAACCAATGTTGCAACGACCTTGA 647  
Qy 217 SerPheIySglIyGluThrCysLeuGluLeuGluGluGlnCysTyrPheTyrIleGlyAsn 236  
Db 648 AGTTTAAAGAAAGAAATATGCTTGAATTACTTGAACAAATGCTATTTTACATTGGAAT 707  
Qy 237 CysGlySapAapIleIleIySglIySglIyGluGluGluGluGluGluGluGluGlu 256

[illegible]

QY	616	ThrCysgluLuewvctVlyvshpVlyvshAspArgCgluVAlPheVlyvshIle	635
Db	1836	ACGtGTAGCGTGAAGTGAAGACGTGAAGACGGTGAAGATTCAAAAAATATA	1895
QY	636	LysAlaSerTrIleIleGluPheGluAAsnshThzAsnLysIleThrThrLeuGlu	655
Db	1896	AAACCTTCATATATATTGAATTTCTTGAAAATATATCAATATAAATPACACACTGGA	1955
QY	656	ArgAnCysPProSerTrPHisThrTrCysAsnArgPheSerProAnCysPProGlyLeu	675
Db	1956	AGAAATGTCCTCTTGCACTACTGTAATGTGAATTTGACCTTAATTTGTCAGAGCTT	2015
QY	676	ThrIySGluAnSerCysPThrLysIleLysLysHisArgGluPProPheTrIyLysArgLys	695
Db	2016	ACGAAAGAGATAGTTGTACAAAATATAGAGAGCATTTGAGCCGTTCTATAAAGAAAG	2075
QY	696	AlaLeuGluAbpAlaLeuLysValGluLeuGlnGlyLysLeuThzAspLysSerLysCys	715
Db	2076	GCTTTGGAGATGCTCTCCAAAGTGAAGCTTCAAGGAAAAATTGACTGTAATATCTAAATGT	2135
QY	716	GluPProAlaLeuLysArgTrCysPThrValAlaGlyAsnAlaIleAsnAlaSerIleSer	735
Db	2136	GAACCTGATTTGAAAAAGATATGTACAGTGAAGCGGAAACGTAAATATATGGCTCATAGT	2195
QY	736	GlyLeuCysLysAlaAsnThrLysAspAsnSerGlyLysSerAspGluAbpAlaArgLys	755
Db	2196	GGCTTATGCAAGCTTAACCAAGAGATACCTCGGAAGAAGTGATGAGATGCTGAGAAAG	2255
QY	756	GluLeuCysGluLysLeuValLysGluValGluGluGlnCysLysValAlaPProThrGlu	775
Db	2256	GAACCTGTGTGAAGAAATTAGTGAAGAAAGATGGAAGACAGTGAACGATTTACCAACGAA	2315
QY	776	LeuGlyGlnPProAlaAlaAspLeuLysAspTrIySerThrTrArgGluGluLeuLysLys	795
Db	2316	TTAGAGCAACCGGAGCTGATTTAAAAAAGATTATAGACATRTAGGAACTTAAGAA	2375
QY	796	ArgAlaGluGluAlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGlu	815
Db	2376	CGTCAAGAGAGCAATGAACCAAGTCCAGTCTTGTTGTCACTCATTAAGAAAAAGAA	2435
QY	816	SerAsnValSerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeu	835
Db	2436	AGTATGATCAATAAAGATTAATGCAAAACAGAGATTAAGATGCGTTCAACCGGACTT	2495
QY	836	GlnAspThrThzLysHisValLysIleLeuArgArgGlyValLysAspValSerValThr	855
Db	2496	CAAGTACCAAAACATGTGAATAATCTACGAGAGAGATTTAAGATGATCTCGTAACA	2555
QY	856	GluLeuGluAlaLysAlaPheAspLeuAlaIleGluValPheGlyArgTrValAspLeu	875
Db	2556	GAATTTAGAGCTAAAGCATTTGATTTGGCAGCGAGAACTAATTTGGAAAGATATGATGATTG	2615
QY	876	LysGluArgCysAsnLysLeuGlnSerAspCysArgLleLysGluAspCysLysAspLeu	895
Db	2616	AAAGAAAGATTAATAATTGCAATCGATTGCGAATTTAAGAGGAGTTGCAAAAGACTTA	2675
QY	896	GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys	915
Db	2676	GAAGAGATATCAAAAAAGATTAATAAGGCTTGTGCAATCTGAAGCCTCTGGAGGTGAAG	2735
QY	916	ProHisGluThrValLtnGlnSerThrThzThrThrThrThrThrThrThzValAla	935
Db	2736	CCGACCGAAACAGACGAAAGTACACGCAACTACAAACAAACAAACAAACACGTTGACC	2795
QY	936	AspProLysAlaThrGluCysLysSerLeuGlnThrThzAspThrTrPValLtnGlnThr	955
Db	2796	GATCGGAAGCAAGCAATGCAATCTCTACAGCAACAAGCAACATGGGTTACACACACA	2855
QY	956	SerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThr	975
Db	2856	TGGACACACACAAACACGCTTCACTATCACTATCACTCACTCACTCACTCACTCACTTACA	2915

QY 976 SerThrArgArgCysLysProThrLysCysThrThrThrGluGluAspAspAlaGlyAsp 995  
DB 2916 TCACGAGCGCATCAACCAACCAAGTGTACGACAGG-----GATGATGACGAAGAC 2969  
QY 996 ValLysProSerGluGluLysLeuArgMetSerGlyTTPanValMetArgGlyValIleVal 1015  
DB 2970 GTGAAGCCAAATGTAAGGCTTGAAGGTGACGGGTGATGTGATGAGAGGGGGTATGTA 3029  
QY 1016 AlameValIleSerPheMetIle 1023  
DB 3030 GCATGCTTATTTGCTTCATCATTT 3053

RESULT 5  
AR438437 3084 bp DNA linear PAT 20-FEB-2004  
LOCUS Sequence 7 from patent US 6664053.  
DEFINITION AR438437  
ACCESSION AR438437.1 GI:42663292  
VERSION  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3084)  
AUTHORS Kovacs,J.A., Huang,S., Maaur,H., Fischer,S.H., GILL,V.J. and Mei,Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG)  
gene of human Pneumocystis carinii  
JOURNAL Patent: US 6664053-A 7 16-DEC-2003;  
FEATURES Location/Qualifiers  
SOURCE 1. 3084  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 2.27e-271 Length: 3084  
Score: 4352.00 Matches: 826  
Percent Similarity: 87.28% Conservative: 80  
Best Local Similarity: 79.58% Mismatches: 106  
Query Match: 80.43% Indels: 26  
DB: Gaps: 6

US-10-654-416-14 (1-1023) x AR438437 (1-3084)

QY 1 MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluGluHis 20  
DB 1 ATGCCGCGGGCGGTCAACGCGCAGCAAAAGTGCACAGATGATGATGAGAGCAT 60  
QY 21 ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluGluCysLysLysLeu 40  
DB 61 GTTTTACCTTGGATTTTAAAAAATGATAGATAGATACAAATGCAAAATGAAAGTTG 120  
QY 41 LysLysLysCysGluGluLeuThrGluAlaLysLeuAsnIleGluGluValHisArgLys 60  
DB 121 GAAACAAATTCGCAAAACATTAACAAATGACAGATTTAAATCCAGAAAAAATTTCAAGAAAA 180  
QY 61 LeuLysGluPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80  
DB 181 TTAAAGATTTCTGTGATACGGAACGAAATGAAATGAAATGTAAGATCTTAAAAACAA 240  
QY 81 IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIle 100  
DB 241 GTCAATCAAAAATGCAATTAATTTCAAGGAAAACTTCAACAGCGCTAGAAAAAAATTT 300  
QY 101 GlnIleIleThrAspLysAspCysLysGluAsnGluGluGluGluCysLeuPheLeuGluGly 120  
DB 301 TCAGAAATTAACAGATGAGATTTGCAAAAAGAAATGAAACAAATCATATTTTGGAGGA 360  
QY 121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys 140  
DB 361 GCATGTCCAAACAGAACTTAAGATGACTGCAATTAATTAAGAAATTAACGTGTTACAAAA 420  
QY 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160  
DB 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160

DB 421 GAAACGAAACAATGTGGCAGAAAGATTTCTTTGAGGGCGCTTCGTGATCTCAATGAA 480  
QY 161 SerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGluSerAsp 180  
DB 481 ACAAAAGCATGTGAAAAAATGCAAAAGTTGGCCCGAAATTTGAAAGAAAGCAT 540  
QY 181 GluLeuThrAsnLeuCysLeuAsnGluLysGluThrCysLysAsnIleLeuIleGluLys 200  
DB 541 GAATTTACGGAGCTTTGCTTTTATCAAAAAACAATGCGTAAGTCTGTGTAACAAAAGA 600  
QY 201 AspLysLysCysGlyThrLeuLysThrAspValSerIleAlaLeuGlySerPheLys--- 219  
DB 601 AAAAGTAAATGTGATCTCTTGAAGAAAGATTGAAACACTTAAAGAAATGATG 660  
QY 220 LysGluThrCysLeuGluLysLeuLeuGluGluCysTyrPheTyrIleGlyAsnCysGlyAsp 239  
DB 661 CGAAGAAAATGTCTACTTATTAATCTTGACCAATGTTACTTTCAAGAGGAACTGGAAGA 720  
QY 240 AspAsp-----IleIleLysCys 245  
DB 721 GACAAATCAAAAGTCATTAACCTAATATATAAGACTGCAAAAGATATGTACAGAGTGT 780  
QY 246 IleGluLeuGluGlyLysCysGluGluGluAsnIleAlaTyrMetProProGlyProAsp 265  
DB 781 GATCAATTTACAGAAAAGTGTGAAAAAATATGTTATATGATCCAGATCCGAT 840  
QY 266 PheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyr 285  
DB 841 TTGCATCAACTAAGCCAGAGCCTACCTAGCAGAGCACTAGGCGCTGAGAGACCTTAT 900  
QY 286 LysLysValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAla 305  
DB 901 AAGAGCGCAGAGAGATGGAATTTTGTGGAAGACACATGTAAAGATGCAACAGCT 960  
QY 306 LeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAspLysGluLysCys 325  
DB 961 TTGTTGGCACTACTT-----CTTAAGAAAAACCTTAAAGAAAGAAATGT 1005  
QY 326 GluGluAlaLeuGluLysSerCysLysAsnProIleGluIleGluAlaLeuGluSerLeu 345  
DB 1006 ATAAAGCCCTTAAAAAACTGGAAACCTCATATAACATGAGGCTTAAAGAAATCTTA 1065  
QY 346 CysLysLysAsnGluLeuSerAsnAspGlyThrLysLysCysGluGluLeuGluAsnAsp 365  
DB 1066 TGTAGGAAAAATTAACCAAGATGATGGAACGAAAAATGTGATGAACCTAAGAAAAAGAT 1125  
QY 366 IleAsnLysThrCysLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspPro 385  
DB 1126 GTTAACAAAACCTGTACAGCTTACATCAACATTTCTTAAAAAACCGTCTTACATTTCA 1185  
QY 386 ThrLysGluAsnAsnGluIleValGlyTTPGluGluLysProThrPheLeuSerAsnGlu 405  
DB 1186 CCGATGAG-----ATTGCGGAATGGGAAAAATTTACCGACATTTCTTAGAGAGA 126  
QY 406 AspCysAlaLysLeuGluSerTyrCysPheTyrPheGluLysLysCysProAspGlyGlu 425  
DB 1237 GATGTGCAAAACATAGATCTTATGCTTTATTTATTAAGAACTTCCAGATGCATA 1296  
QY 426 AsnAlaCysLysAsnIleArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsn 445  
DB 1297 GAAGCTTGTAATGATGAGAGCGCGCTGTACAAAGAGAGCGCTGTATGACGCGGCAAC 1356  
QY 446 LysValLeuGluGluAsnMetArgGlyMetLeuHisGlySerAsnLysSerTyrLeuGlu 465  
DB 1357 AGCTGTGTGCAAAAATTAAGCGGGGTATTAAGTGTTCAAATCAAGGTGCTTAAG 1416  
QY 466 LysPheGluGluGluLeuValLysValCysGluLysLysLysGluAsnLysGlySer 485  
DB 1417 GAGTTTCAACAAAGATTAAGTAAGATGTATGAGAGCTA---AAAGAAAAATTAAGAGAT 1473  
QY 486 PheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAlaAlaArgLeuLeu 505  
DB 1474 TTCCCAACGATGAATAATATTTGTTCTGTGTGTACAGCAGCAAAAGCTGACGATTACTT 1533

QY 506 ThrHISaPLeuAysMeLysThrIlePheLeuAtrGInGInLeuAspGInuLysArgAsp 525  
 DB 1534 ACACACGATTCATCAAAATAGAGGTTACTCTTTTACCAACAATGGAATGCAAAAGAGAT 1593  
 QY 526 PheProTArAspLysAsnCySLysGluLeuGlyArgLysCySLysAspLeuGluYqLysAsp 545  
 DB 1554 TTTCGACGAGTAAGAGACTGACAGAGACTAGGGAAATGCGCAAGATTTAGAAAGGAT 1653  
 QY 546 SerLysGluIleThrTrpProCysHisThrLeuGluGInGInCysAsnArgLeuGluYThr 565  
 DB 1654 TCAAAAGAAATTACATGGCCATGTCATACCTGAGAGCAATGCAATCGCTGGGCACT 1713  
 QY 566 ThrGluIleLeuLysGInValLeuLeuAspGluHisLysAspThrLeuLysAspGInu 585  
 DB 1714 ACAGAAATTTAAACAGGTTTATTTAGATGAAACAAAGATACCTTAAAGACCAAGAA 1773  
 QY 586 SerCysValLysTrpLeuLysGluLysCysAsnLysTrpSerArgArgGluYAspAspArg 605  
 DB 1774 AGTTGTAAATACCTTAAGAAAGATGTAAATAGGTCTGAAAGAGAGATGACCGT 1833  
 QY 606 PheSerPheValCysValPheGInAsnA1aThrCysGluLeuMetValLysAspValLys 625  
 DB 1834 TTCTCTTTTGTATGTGTTTCCAAACCGCTACGTGTGAGCTGATGTTAAAGACGTGAA 1893  
 QY 626 AspArgCysGluValPheLysLysAsnIleLysAlaSerTyrIleIleGluPheLeuGlu 645  
 DB 1894 GACAGGTGTGAAGTATTCAAAAAAATATATAAGCTCATATATATATGGAATTTCTTAA 1953  
 QY 646 AsnAsnThrAsnLysIleThrThrLeuGluArgAsnCysProSerTrpHisThrTyrCys 665  
 DB 1954 AATAATACAAATATAATACAACTCGAAGAAATTTGCTCTTGGATACGTAATTC 2013  
 QY 666 AsnArgPheSerProAsnCySProGluLeuThrLysGluAsnSerCysThrLysIleLys 685  
 DB 2014 AATAGATTTTGCCTTAATGTGTCAAGCGCTTACGAAAGAAATGTGTAACAAAATCAAG 2073  
 QY 686 LysHisAlaArgGluProPheTyrLysArgLysAlaLeuGluAspAlaLeuLysValGluLeu 705  
 DB 2074 AAGCATTTGTAGCCCTTCATATAAGAAAGAGCCCTTGAAAGATGCTCAAAAGTAAAGCTT 2133  
 QY 706 GInGluLysLeuThrAspLysSerLysCysGluProAlaLeuLysArgTyrCysThrVal 725  
 DB 2134 CAAGGAAATTTGACTGATTAATCTTAATGTGAACCTGCAATGACAGATTTGTACAGTA 2193  
 QY 726 AlaGluAsnValAsnAsnAlaSerIleSerGluLeuCySLysAlaAsnThrLysAspAsn 745  
 DB 2194 GCGGGAACGTAAATATATGCGTCATCACTGTGCTTATGCAAGCTAACCCAGAGATTAAC 2253  
 QY 746 SerGluLysSerAspGluAspAlaArgLysGluLeuCySGluLysLeuValLysGluVal 765  
 DB 2254 TCTGAAAGAGATGAGAGATGCTGAAAGAAAGAACTCTGAAATACTGATGAATAAGATG 2313  
 QY 766 GInGluGInCysValAlaLeuProThrGluLeuGInProAlaAlaAspLeuLysLys 785  
 DB 2314 GAAAGACGTCGAAACGATTAACCAAGATTAAGCAACCGGACGCTGATTAATAAA 2373  
 QY 786 AspTyrLysThrTyrGluGluLeuLysArgAlaGluGluAlaMetAsnLysSerSer 805  
 DB 2374 GATTATTAAGACATATAGAGAACTTAAGAAACGTGACAGAGAAACATGAACAAGTCCAGT 2433  
 QY 806 LeuValLeuSerLeuIleLysLysAsnGluSerAsnValSerLysSerAsnSerLysAsn 825  
 DB 2434 CTGTGTTTGTCACTCATTAAGAAAAACAAAGTATGTAACAAAAGTATGCAAAAAC 2493  
 QY 826 LysAspLysAsnAlaValSerAsnGluLeuGInAspThrLysHisValLysIleLeu 845  
 DB 2494 AAGGATTAAGATGCCGTTCAACGCACTTCAAGATTAACCAAAAACATGTGAAATTACTCA 2553  
 QY 846 ArgArgGluValLysAspValSerValThrGluLeuGluAlaLysAlaPheAspLeuAla 865  
 DB 2554 CGGAGAGGAGTAAAGAGATGATCCGTAACAGAAATTAAGAACTTAAGCAATTTGATTTGGCA 2613

QY 866 AlaGluValPheGluYArgTyrValAspLeuLysGluArgCysAsnLysLeuGluSerAsp 885  
 DB 2614 GCAGAGATATTGGAAAGATATGATGATTGGAAGAAAGATGTAAATTAATGCAATCAAT 2673  
 QY 886 CysArgIleLysGluAspCySLysAspLeuGluGluValCysLysLysIleAsnLysAla 905  
 DB 2674 TGCAGAAATTAAGACGATTCGAAAGACTTGAAGAAAGATTAAGATTAATTAAGGCT 2733  
 QY 906 CysArgAsnLeuLysProLeuGluValLysProHisGluThrValThrGluSerThrThr 925  
 DB 2734 TGTGCAATTCGAAGCCCTCTGAGGTGTAAGCCGACGAAACAGTGCAGAGTAACAAG 2793  
 QY 926 ThrThrThrThrThrThrThrThrValAlaAspProLysIleThrGluCysLysSerLeu 945  
 DB 2794 ACAACTACAAACAAACAAACAAACCGTTGCCATCCGAAGGAGAGAAATGCAAAATCCTTA 2853  
 QY 946 GlnThrThrAspThrTrpValThrGluThrSerThrHisThrSerThrSerThrIleThr 965  
 DB 2854 CAGACAAAGACACATGGGTTACACAGACATCGACACACACACACACACAGCTTACTACCA 2913  
 QY 966 SerThrIleThrSerLysIleThrLeuThrSerThrArgArgCysLysProThrLysCys 985  
 DB 2914 TCTACCATCATCAATCAAAATATACATTGACATCAACGAGCGATGCAAAACCAACAGTGT 2973  
 QY 986 ThrThrGluGluGluAspAspAlaGluAspValLysProSerGluGluLysArgMetSer 1005  
 DB 2974 ACGACAGGG-----GATGATGTCAGAAAGCTGAAAGCCAACTGAAGGCTTAAAGGTTGAGC 3027  
 QY 1006 GlyTrpAsnValMetCysArgLysValIleValAlaMetValIleSerPheMetIle 1023  
 DB 3028 GGGTGAATGTGATGAGGGGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 3081  
 RESULT 6  
 AF033209 3083 bp DNA linear PLN 10-SEP-1998  
 LOCUS  
 DEFINITION Pneumocystis carinii f. sp. hominis clone HumsG14 major surface glycoprotein (MSG) gene, partial cds.  
 ACCESSION AF033209  
 VERSION AF033209.1 GI:3560514  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 . Pneumocystis jirovecii  
 Pneumocystis jirovecii  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.  
 REFERENCE  
 AUTHORS Garbe, T.R. and Stirling, J.R.  
 TITLE Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii  
 JOURNAL Infect. Immun. 62 (8), 3092-3101 (1994)  
 MEDLINE 94314421  
 PUBMED 7518806  
 REFERENCE  
 AUTHORS Mei, Q., Turner, R.E., Sorial, V., Klivington, D., Angus, C.W. and Kovacs, J.A.  
 TITLE Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region  
 JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)  
 MEDLINE 98380374  
 PUBMED 9712777  
 REFERENCE  
 AUTHORS Mei, Q., Turner, R., Sorial, V., Klivington, D., Angus, C.W. and Kovacs, J.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA  
 FEATURES  
 source  
 1..3083  
 /organism="Pneumocystis jirovecii"  
 /mol\_type="genomic DNA"  
 /specific\_host="Homo sapiens"  
 /sub\_features="hominis"



OY	507	H1sApRLeuAaGMeLysThrI1LePheLeuAArgGInGInLeuAArgGInLysAArgPhe	526
Db	1536	CACGATCATCAAAATGAAGGTTACCTTTTATGACAAACAATTGGATCAAAAGAGAAATTT	15959
OY	527	ProThrAspLysAsnCyLeuG1uLeuG1yAArgLysCyGInAspLeuG1yG1uAspSer	546
Db	1596	CCGACAGATTAAAGACTGCAGAGAACTACGAGAAAAAATCCCAAGATTTCAGAAAGATTCA	16555
OY	547	LysG1uL1ethrTPProCyAsH1ethrLeuGInGInCyAsAsnArgLeuG1yThrThr	566
Db	1656	AAAGAAATTACATGGCCATGTCTATCACTGGAGCAGCATCAATCGCTTGGGACTCA	17151
OY	567	GlulLeuLysGInValLeuLeuAsnG1uH1sLysAspThrLeuLysAspG1uGlnSer	586
Db	1716	GAATATTTAAAGCAGGTTTATTTGATGATACCAAAAGATCTTGAAGACCAAGAAAT	17757
OY	587	CysValLysTyrLeuLysG1uLysCyAsAsnLysTPSerAArgG1yAspAspArgPhe	606
Db	1776	TGTGTAAATACCTAAAGAAAGAAAGCTAAATATGTCTTAAGAGAGATGACCGTTTC	18353
OY	607	SerPheValCyValPheGlnAsnA1aThrCyGInLeuMetValLysAspValLysAsp	626
Db	1836	TCTTTTGTATGTGTTTCCAAACCCCTACGTGTGAGCTGATGGTAAAGACGTGAAGAC	18959
OY	627	ArgCyGInValPheLysLysAsnL1eLysA1aSerTyrI1eGlnPheLeuGlnAsn	646
Db	1896	AGGTGTGAAGTATCTAAABAAAATTAAGAAGTTATATATATTAATGAATTTCTGAAT	19555
OY	647	AsnThrAsnLysI1eThrTrLeuGlnAArgAsnCyPProSerTyrPheI1eThrCyAsAn	666
Db	1956	AATACAAATTAATTAACAACACTGGAAAGAAATGTCCCTCTGGCATACGTAATTGCAT	20151
OY	667	ArgPheSerProAsnCyPProG1yLeuThrLysG1uAsnSerCyPThLysI1eLysLys	686
Db	2016	AGATTTTCAACCTAATGTCCAGGCTCTACGAAAGAAATAGTTGTACAAAATCAAGAA	20757
OY	687	H1sArgLysPProPheTyrLysAArgLysA1aLeuG1uAspA1aLeuLysValGlnLeuGln	706
Db	2076	CATGTGTAGCCGTTCTATTAAGAAAGAGCCTTGAGAAAGTGCCTCAAGTAAGAGCTTCA	21353
OY	707	G1yLysLeuThrAspLysSerLysCyGInProA1aLeuLysAArgTyrCyAsThrValA1a	726
Db	2136	GGAAATTTGACGTAAATCTTAATCTGAACCTGCATTTGAACAGATATTGTACAGTACG	21959
OY	727	G1yAsnValAsnAsnA1aSerI1eSerG1yLeuCyLysA1aAsnThrLysAspAsnSer	746
Db	2196	GGAAACGTAAATTAAGCGTCAATCAGTGGCTTATCCAAAGCTAACCAAGATAACTCT	22555
OY	747	G1yLysSerAspG1uAspA1aArgLysG1uLeuCyGInLysValLysG1uValG1u	766
Db	2256	GGAAAGATGATGACGATGCTTAAGAAAGAACTCTGTGAAATTCGTAAAGAAAGTGA	23151
OY	767	GlulGInCyLysA1aLeuProThrGlnLeuGInGInProA1aAspLeuLysLysAsp	786
Db	2316	GAACAGTCCAAAGCATTTCCAAACAAGATTAGGACAAACCGGACGTGATCTAAAAAAGAT	23757
OY	787	TyrLysThrTyrGlnGlnLeuLysAArgA1aGlnGlnA1aMetAsnLysSerSerIeu	806
Db	2376	TATAAGACATATGAGGAATTAAGAAAGCTGACAGGAAGCAATGAAACAGTCCAGTCTT	24353
OY	807	ValLeuSerLeuI1eLysLysAsnG1uSerAsnValI1eSerLysSerAspSerLysAsnLys	826
Db	2436	GTTTGTCTCATTAAGAAAGAAAGAAAGTAAATGTATCAAAAAGTAATAGCAAAAACAG	24959
OY	827	AspLysAsnA1aValSerAsnG1yLeuGInAspThrThrLysPheValLysI1eLeuArg	846
Db	2496	GATTAAGATGCCGTTTCAAGGCACTTCAAGTATCACAAAAACATGTAAAAATACTACG	25555
OY	847	ArgG1yValLysAspValSerValThrGlnLeuGlnA1aLysA1aPheAspLeuA1aA	866
Db	2556	AGAGGAGTTAAAGATGATCCGTAAACAGATTAAGAAAGCTTAAGCACTTTGATTGGACGA	26151

OY		867	GluVal PheGlyAArgYrValAspLeuLYGIuArgCYasnuLYseuGIuseArSPCyS	866
Dd		2616	GAAGTATTGGAAATATGTGATTTCAAGAAAAGATGTAATAATTGCATCATGATTC	2675
OY		887	ArgIleLygLIuaPcYelysaPLeuGIuValCylLySylylleanLYsAlaCys	906
Dd		2676	AGAATTAAAGCGAGTCCAAAGCTTGAAAGAGTAATGCAAAAAGATTAATTAAGGCTTGT	2735
OY		907	ArgasnLeuLYSProlauGIuValLYSPROHIGLuthrValThrgIuseTrhThTh	926
Dd		2736	CGCAATCTGAAGCCTCTGGAGGTGAMCCCCGACGAAACAATGACAGNAGGTACMAAGCAC	2795
OY		927	ThrThrThrThrThrThrThrValAlaAspProLYsAlaThrgIusLYSeSerLeuGIu	946
Dd		2796	ACTCAACAACAACAACACCGTTGCCGATCCGAAGCAAACGGAATCAATCTTAATC	2855
OY		947	ThrThrAspThrTPvalThrgInThSerThHiThiSerThrserThrlieThSer	966
Dd		2856	ACAAcAGAcATGGGTTCACACACATTCACACACACAAACGCTTACTATTCACATCT	2915
OY		967	ThrIleThSerLYsIleThrLeuThSerThArghArgCYeLYSProThrLYSCyeThr	986
Dd		2916	ACCATCACAATAAAAATACATTGACATCAACAGGCGATGCCAACCAACCAATGATACG	2975
OY		987	ThrgIyGLIuaAPaPaIagLYaSPvaIIyaSPSeSerGIuLYeuArqMeSPeSergLY	1006
Dd		2976	ACAGGG-----GATGATGCGAGAAGCGTAGAGCCAAGTAAAGCTTGAGCGGTACGCGG	3029
OY		1007	TrypanValMetArgLYValIIleValAlaMerValIleSerPhemeticIle	1023
Dd		3030	TGSAATGTGATGAGGGGGGTGATGTAGTCANATGGTTATTTCGTTCAATGATT	3080
RESULT 7				
LOCUS	AR438436	3090 bp.	DNA	linear PAT 20-FEB-2004
DEFINITION	Sequence 5 from patent US 6664053.			
ACCESSION	AR438436			
VERSION	AR438436.1	GI:42663291		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 3090)			
AUTHORS	Kovacs,J.A., Huang,S., Maeur,H., Fiescher,S.H., Gill,V.J. and Mei,Q.			
TITLE	Identification of a region of the major surface glycoprotein (MSG) gene of human Pneumocystis carinii			
JOURNAL	Patent: US 6664053-A 5 16-DEC-2003;			
FEATURES	Location/Qualifiers			
source	1..3090			
	/organism="unknown"			
ORIGIN	/mol_type="genomic DNA"			
Alignment Scores:				
Pred. No.:	1..666-237	Length:	3090	
Score:	3827.00	Matches:	728	
Percent Similarity:	81.54%	Conservative:	120	
Best Local Similarity:	70.00%	Mismatches:	164	
Query Match:	70.73%	Indels:	28	
DB:	6	Gaps:	8	
US-10-654-416-14 (1-1023) x AR438436 (1-3090)				
OY	1	MetaLaarGaLaVaLIySaRgGlMaLaIaGLYThrGlnAnserIleaRPGIuLnHIS	20	
Dd	1	ATGGCGCGGGCGGTCAAGCGCGGCAAAAGGTGCACAGAAATGACATTGATGAGAGCAT	60	
OY	21	ValleuMalaenLIleLeuLYSGIuaSPcYLEuSerGIuNGIuGIuSYLySLyLYeU	40	
Dd	61	GTTTAGCTTGATTTTAAAAAAAATGATGATGAAATATCAAAATCCAAACTAACTTG	120	
OY	41	LYsyLYrCYeGIuNGIuLeuThGIuAlaLYsLeuAnIIleGIuValIHIsaRyLS	60	





Qy	764	GIUVALGIUGIUGINCYSVLVALELEMPROTHRGIUENUGIINPROIALALEAPREU	783
Db	2314	GAGGTGGACGCAATGCAGATATTACCAACGATTAACAGAGCTGGAAAAAGTCTA	2373
Qy	784	LYSLYSAPRYLYLSTHYTRGLUGIUENULYSLVATGALAGIUGIUALAMEFASNLVS	803
Db	2374	AAAAAAGATGTTAAGACATATAGAGAACTTAAGGAAAGGGCAAAAAAGCAATGACAG	2433
Qy	804	SerSerLeuValLeuSerLeuIleLysLysAngIuSerAenValSerLysSerAenSer	823
Db	2434	TCCAGCCTTGTTTATTACTCTGTTAAGAAAAACGAAGTAATATACCGAAAAATPATAGC	2493
Qy	824	LYSAnLYSAPRYLSAenAlValSerAenGIUENGIuAenPRThThLysHISValLys	843
Db	2494	AAAAACAAGGATTAAGAAATGCTGTTCAACGCACTCAAGATACCACAAAATATATGAAA	2553
Qy	844	IleLeuATGATGAGIValLysAAsPValSerValThrGIuLeuGIUALALysAlAPheAsP	863
Db	2554	ATTACTCAAAAGAGAGTTAAGGAGCACTTGTAACGAATCTGAAGCCAAAGCAATTGAT	2613
Qy	864	LeuAlaAlaGIUVALPheGIYARGTyrValAsPLeuLysGIUARGCySAsnLysLeuGIU	883
Db	2614	TTGGCAGACGAGAGTGTGGAAGATATGTAACTTGAAAGAAAAATGGAAGAAATTGACT	2673
Qy	884	SerSerCySAsnGIleLysGIUAsPLeuSAsPLeuGIUGIUALCySALysLysIleAsn	903
Db	2674	TCCGATTCGGGATTTAAAGACGATGGCGATGTTAAAAAGAGTGTGGAAGAGATTGAG	2733
Qy	904	LYSAlaCySAsnLeuLysProLeuGIUVALYSPROHISGLThThValThrGIUSeR	923
Db	2734	AAGACATGTCCAGATCTGAACCTCTGAGGTGAAGTCCGATGAATGTCAACGAAAGC	2793
Qy	924	ThrThrThrThrThrThrThrThrThrThrValAlaAsPProLYSAlaThrGIUCyLys	943
Db	2794	ACAACGACGACCAACAACAACAACGACCGTTACCGATCCGAAGCAACAGAAATGCAAA	2853
Qy	944	SerLeuGIuThrThrThrAsPThrTyrValLhrGIuThrSerThrThrHISerThrSerThr	963
Db	2854	TCCTTTACGACAACGATACATGAGGTTCACACAGATGACACACACAAGACCGTTAC	2913
Qy	964	IleThrSerThrIleThrSerLysIleThrLeuThrSerThrAsnArgCySAsPProThr	983
Db	2914	ATCACATTTACATCATCATTCAAAAATTAACATTGACATACAGAGCGCATGCMAACCAAC	2973
Qy	984	LysCySThrThrThrGIUGIUALSAsPAsnAlaGIYAsPValLysProSerGIUGILeuArg	1003
Db	2974	AAGGTACGACAGAGG-----GATGAACAGAGAGACGTAACCGAGTGAAGATTGAAG	3027
Qy	1004	MetSerGIYTPAsnValMetArgGIYValIleValAlaMetValIleSerPheMetIle	1023
Db	3028	ATGATGTGGTGGAGCGGTATGATGAGGGGGTGAATGATGAAGAATGTTATTCGTTATGATT	3087
RESULT 8			
LOCUS	AF033208	3089 bp	DNA linear
DEFINITION	Pneumocystis carinii f. sp. hominis clone HUMS011 major surface glycoprotein (MSG) gene, partial cds.		
ACCESSION	AF033208		
VERSION	AF033208.1	GI:3560512	
KEYWORDS	Pneumocystis jirovecii		
SOURCE	Pneumocystis jirovecii		
ORGANISM	Pneumocystis jirovecii		
REFERENCE	1 (bases 1 to 3089)		
AUTHORS	Garde, T.R. and Stringer, J.R.		
TITLE	Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii		
JOURNAL	Infect. Immun. 62 (8), 3092-3101 (1994)		
MEDLINE	94314421		
PUBMED	7518806		
REFERENCE	2 (bases 1 to 3089)		
AUTHORS	Mei, Q., Turner, R.E., Social'V., Klivington, D., Angus, C.W. and		

**TITLE** Kovacs,J.A.  
 Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region  
**JOURNAL** Infect. Immun. 66 (9), 4268-4273 (1998)  
**MEDLINE** 98380374  
**PUBMED** 9712777  
**REFERENCE** 3 (bases 1 to 3089)  
**AUTHORS** Wei,Q., Turner,R., Social,V., Livingston,D., Angus,C.W. and Kovacs,J.A.  
**TITLE** Direct Submision  
 Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA  
**FEATURES**  
 source Location/Qualifiers  
 1..3089  
 /organism="Pneumocystis jirovecii"  
 /mol\_type="genomic DNA"  
 /specific\_host="Homo sapiens"  
 /sub\_species="hominis"  
 /db\_xref="taxon:42068"  
 /clone="HUMSG11"  
 /note="derived from HIV-infected human with P. carinii pneumonia"  
 <1..3089  
 /gene="MSG"  
 <1..3089  
 /gene="MSG"  
 /note="surface antigen"  
 /codon\_start=3  
 /product="major surface glycoprotein"  
 /protein\_id="AAC34971.1"  
 /db\_xref="GI:3560513"  
 /translation="ARAVKRAKGAQNSIDSEHVALLIKKNGLEDTCCKTXLEYCK  
 TLINAGINPEKVEHLEKIDPCDNGKRNKECODLRKNYOKCIFKQGLQTAQKIKSEL  
 TDDCKNKEOQGLFEGACPTLKDCCNLRNNCYOKERNVAEVLRLRLDGLNET  
 KTECKLLEKCYPTLRBSDELTELCLYKQTKVSLVTKGSKCDTLEKVEALKKNE  
 LSKCKLLEKCYFHRNGCEGDKSKKRNPKCKEYVEPCDELAELCKKGENIYVNP  
 GSFDPFKPEPTLAIEDIGLELYKRAEBDI FGRGHVRAVATLALLLTKTLKKEC  
 IKALKKCNENPHEHEALENLCKENKSSDPTKCDLEKDVNTCTSLTILKRLY  
 ISPDGLAEWKEPLTFLSDEDCALLESYCFYKETCPDVKACANVPAACYKRLDARA  
 NVSLQKNRGLHLSGKKDMLKKRQELAKYCEKLKGNKGSFNSDELFLVCIOPAKAR  
 LILTHHQMRVYIFLRQLDOKRDPPTDKCELEGRKODLKGSKELTWPQHTLEOCN  
 RLGLTILKQILIDHDKDTLKHENCAKYLKRAKCHSRRRGDRFSFVCFQVATCEL  
 MVDVODRCKIFEENQAQIDINDLSLKKNOIKASANI CPSMRYCDRLFVPCPDLLK  
 GKTFCKNKKCYCEPFYKRYVLEINDLYELKGNLSNTKCPALERYCTVLPNDVMSI  
 SSKCNDESCTKTKADKNVRKYLCLKLYEVEQOQKVLPTLELLEKSLKADVKTYE  
 ELKEPRKANKRSILVLSLVKKNESNTSKNSKNDKNVSNGLDPTTKYVKILRGV  
 KHALVTRSEAKADPLAAEVGRVYDLKEKCEKLTSPGCIKDDCDGLKYNCGKLEKCH  
 DLKPLEKSHIEIVTESITTTTTTTTTTVDTPKATCEKSLQTTDVTWVOTSHISTIT  
 STITSKITLSTSRCKPTKCTTDEAGDVKPSBGLKMSGWSVMRGVIAVMVIFSMI"  
**ORIGIN**  
 Alignment Scores:  
 Fred. NO.:  
 Score: 3.48e-237 Length: 3089  
 Percent Similarity: 3822.00 Matches: 727  
 Best Local Similarity: 81.52% Conservative: 120  
 Query Match: 69.97% Mismatches: 164  
 70.63% Indels: 28  
 Gaps: 8  
 US-10-654-416-14 (1-1023) x AF0333208 (1-3089)  
 Oy 2 AAlarGAlaVallybSaTcGAlnaAlaAGlYThrClnaSserIleaPpGluGluHleVal 21  
 Db 3 GCGCGGGCGGCTCAAGCGGGCGGCAAAAGGTGCACAGAAATGCAATTGATGAGAGCATGTT 62  
 Oy 22 LeuAlaLeuIleuLeuLybGluApsGlyLeuSerGluGluGluGluGluGluGluGluGlu 41  
 Db 63 TTACCTTTGATTTTAAAAAATGATTTTGAAGATGACAAAGACAAAGCTAAGTTGAA 122  
 Oy 42 LybTTCyGsgInGluLeuThrGluAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 61

Db 123 GAATATTGCAAAACATTAAACAAATGACAGATTAAATCCAGAAAAAGTTCCAGAAAAATTAA 182  
 Qy LygIlyPheCysGluAerGlyValAlaSerThrIysCysLysGluLeuValAlaSerIle 81  
 Db 183 AAGATTTCGTGTATACGGGAACGAAATGAAATGTCATGAAATCTTAAACAAAGTTC 242  
 Qy 82 GlnLysLysCysThrIleLeuGlyValLeuLysGluValAlaIleLysLysIleGln 101  
 Db 243 AATCAAAAATGCAATTAATTTCAGGAAAACTTCAACAGCTGCGGAAAAAAATTTCA 302  
 Qy 102 IleIleThrAspLysAspCysLysGluAerGluAerGluAerGluPheLeuGluVal 121  
 Db 303 GAATTAACAGATGAGATTGCAAAAAGATGACAAACAAATGCTTATTTTGAGAGGAGCA 362  
 Qy 122 CysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys 141  
 Db 363 TGTCCAAACAGAACTTAAGATGACGCAATTAATTAAGAAATTAATCTGTATCAAAAAAGAA 422  
 Qy 142 ArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGlySer 161  
 Db 423 CGGAACAATGCGCAGAAAGAGTTCTTTGAGGCGCTTCCTGTGATCTCAATGAAACA 482  
 Qy 162 ValIleCysGluLysLysLysGluIleCysAspValMetGlyArgGluSerAspGlu 181  
 Db 483 AAGACATGTGAAAAAAGCTGAAAGAGTTGCCGAAATTAAGAAAGAGGAGATGAA 542  
 Qy 182 LeuThrAsnLysCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLysAsp 201  
 Db 543 TTAACGAGCTTGTCTTATCAAAAAACAAATGCGTAAGCTTTGTAAACAAAGGAAAA 602  
 Qy 202 LysLysCysGlyThrLeuLysThrAspValSerAlaIleGluLysSerPheLys---Lys 220  
 Db 603 AGTAATAGTGTATCTCTTGAAGAAAGAGTTGAAGAAGACCTTAAGAAGAAATTCGGA 662  
 Qy 221 GlnThrCysLeuGluLeuLeuGluGlnCysTyrPheTyrIleGlyAsnCysGlyAspAsp 240  
 Db 663 GAAAAAAGTCTACTATTACTTGACCAATGTTACTTTCACAGAGGAACTGTGAAGAGAGAC 722  
 Qy 241 Asp-----IleIleLysCysIle 246  
 Db 723 AATCAAAAGTCAATTAACCTTAATTAATTAAGACGTCAAAAGAAATGTAACAGAGTGTGAT 782  
 Qy 247 GlnLeuGlyGlyLysCysGlnGluGlnAsnIleAlaTyrMetCProGlyProAspPhe 266  
 Db 783 GAATTACAGAAAAGCTGTGGAAGAAATATGTTTATATGATCCAGATCCGATTC 842  
 Qy 267 AspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGlnPheTyrLys 286  
 Db 843 GATCAACTAAGCCAGAGCCTTAAGCTAGCAGAGACATAGGGCTGGAAGACCTTATTAAG 902  
 Qy 287 LysValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAlaLeu 306  
 Db 903 AGGCAAGAAAGATGAAATTTTGTGGAAGACAACTGTMAAGATGACAAAGCTTTG 962  
 Qy 307 LeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAspLysGluLysCysGlu 326  
 Db 963 TTGGCACTACTT-----CTTAAGAAAACCTTAAGAAAAGAAAGATGATATA 1007  
 Qy 327 GlnAlaLeuGlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluSerLeuCys 346  
 Db 1008 AAGCCCTTAATAAAAAAACCCTCATGAACATGAGGCTTGAAGAAATCTATGAT 1067  
 Qy 347 LysLysAsnGlyLysSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAspIle 366  
 Db 1068 AAGGAAAAATTAACCAAGTATGATGAAACGAAAAATGTGATGAACATAAGAAAAAGATTT 1127  
 Qy 367 AsnLysThrCysLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspProThr 386  
 Db 1128 AACAAAACCTTACAGATCTTACATCAACATTTCTTAATAAACCGCTTTTACATTCACCT 1187  
 Qy 387 LysGlyAsnAsnGlnIleValGlyTyrGluGlyLeuProThrPheLeuSerAsnGluAsp 406  
 Db 1188 GATGGA-----ATTGCGGAATGGGAAAAATTAACCGACATTTCTTAAGATGAAGAT 1238

Qy 407 CysAlaLysLeuGlnSerTyrCysPheTyrPheGluLysCysProAspGlyGluAsn 426  
 Db 1239 TGTCCAAATCAAGAAATCTTATGCTTTTATTAATAAGAACTTGTCCAGATGTCAAGAA 1298  
 Qy 427 AlaCysLysAsnIleArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsnLys 446  
 Db 1299 GCTTGATGAATGAGGAGCGCTGTTTATTAAGAGAGCGCTTGATGACGCGCAACAGT 1358  
 Qy 447 ValLeuGlnGluAsnMetArgLysMetLeuHisGlySerAsnLysSerTrpLeuGluLys 466  
 Db 1359 GTGTGCAAAAAATTAAGCGAGGATTAATGATGCTCAAAATTAAGATGAGCTTAAGAAA 1418  
 Qy 467 PheGlnGlnGluLeuValLysValCysGlyLysLeuLysGluAsnLysGlySerPhe 486  
 Db 1419 TTTCAACAGAAATTAAGCAAAAGTATGAGAAATG---AAGCAAAATTAAGAAAGTTTC 1475  
 Qy 487 SerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysValAlaArgLeuLeuThr 506  
 Db 1476 TCGAAACATGAAATGTTGTTCTGTGTATCAACCAACCAAGCAGCAGCATTTACTTACA 1535  
 Qy 507 HisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGluLysArgAspPhe 526  
 Db 1536 CATCAACCATCAAAATGAGATTAATCTTTTACGACCAACACTGATCAAAAGAGAGATTT 1595  
 Qy 527 ProThrAspLysAsnCysLysGluLeuGlyValArgLysCysGlnAspLeuGlyLysAspSer 546  
 Db 1596 CCGACAGATTAAGACTCAAGAAATTAAGGAGAAATCCCAAGTTTAAGAAAGATTTCA 1655  
 Qy 547 LysGluIleThrTrpProCysHisThrLeuGlnGlnCysAsnAsnArgLeuGlyThrThr 566  
 Db 1656 AAGCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1715  
 Qy 567 GlnIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeuLysAspGlnLysSer 586  
 Db 1716 GAAATTTTAAAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1775  
 Qy 587 CysValLysTyrLeuLysGluLysCysAsnLysTyrSerArgArgGlyAspAspArgPhe 606  
 Db 1776 TGTCCAAATTAATTAAGAAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTA 1835  
 Qy 607 SerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValLysAspValLysAsp 626  
 Db 1836 TCTTTGATGATGATTTCCAAAGCTACATGTGAGCTGATGATGATGATGATGATGATGAT 1895  
 Qy 627 ArgCysGluValPheLysLysAsnIleLysAlaSerTyrIleIleGluPheLeuGluAsn 646  
 Db 1896 AGGCAAAATTAATTAAGAAATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTA 1955  
 Qy 647 AsnThrAsnLysIleThrThrLeuGluArgAsnCysProSerTrpHisThrTyrCysAsn 666  
 Db 1956 AATCAAAATTAAGCAAGATGACGAGCAAAATTTTGTCTCTCATGCGATCCATCTGCGAT 2015  
 Qy 667 ArgPheSerProAsnCysProGlyLeuThrLys---GluAsnSerCysThrLysIleLys 685  
 Db 2016 AGATTTTACCAATGATGCTGATCTTAAGAAAGAAAACTTCTGTCAAAATCTTAAA 2075  
 Qy 686 LysHisArgGluProPheTyrLysArgLysAlaLeuGluAsnAlaLeuLysValGluLeu 705  
 Db 2076 AATATTTGCGAACATTTCAAAAGAAAGTTTAAAGATGCTCTTAAAGTGAAGCTT 2135  
 Qy 706 GlnGlyLysLeuThrAspLysSerLysCysGluProAlaLeuLysArgTyrCysThrVal 725  
 Db 2136 CGAGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2195  
 Qy 726 AlaGlyAsnValAsnAsnAlaSerIleSerCylLeuCysLysAlaAsnThrLysAspAsn 745  
 Db 2196 TTGAAGACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2255  
 Qy 746 SerGlyLysSerAsp---GluAspAlaArgLysGluLeuCysGluLysLeuValLysGlu 764  
 Db 2256 ACTTAAAGGCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2315

Qy		765	ValGluGluGluGlnCysLeuValAlaLeuProThrGluLeuGluGlnProAlaAlaAspLeuLys	784
Db		2316	GTGAAGACGCATCGCAAAAGTATTATTCACAACAATAAATTAACAGAGCTGGAAAAAAGTTCTAAA	2375
Qy		785	LysAspTrpXylSerThrTrpYrGluGluMetLeuLysLysArgAlaGluGluAlaMetAsnLysSer	804
Db		2376	AAAGTGTTAAGACTATNGAGGAATTAAAGAAAGGCGCAAAAAAGCAATGACAAAGCC	2435
Qy		805	SerLeuValLeuSerLeuIleLysLysAsnGluSerAsnValSerLysSerAsnSerLys	824
Db		2436	AGCCTTGTTTATCATCTGTTAACGAAAAAGAAAGTAATATCATGCAAAAAATTAATGACAA	2495
Qy		825	AsnLysAspLysAsnAlaValSerAsnGlyLeuGluAsnProThrTrpLysHisValLysLe	844
Db		2496	AACAAGGATTAAGAATGTCGTTTTCAAACGCACTTCAGATACCACAAAAATGTAAGTAATA	2555
Qy		845	LeuArgArgGlyValAllyLysAspValSerValThrgLueGluAlaLysAlaPheAspLeu	864
Db		2556	CTACGAGAGAGAGTTTAAGSAGGCACTTTAACAGAAATCGAAGCCAGAGCAATTTGATTTC	2615
Qy		865	AlaAlaGluValPheGlyArgTyrrValAspLeuLysGluCybAsnLysLeuGluSer	884
Db		2616	GCAGACAGAAAGTCTTGGAAGATATGTACACTTGAAGAAAGAAATGAGAAATTTGACTTCG	2675
Qy		885	AspCysArgGlyIleLysGluAspCysLysAspLeuGluGluValCybLysLysIleAsnLys	904
Db		2676	GATTGCCGGATTTAAAGACGATTTGCCATGCTTTAAAGAAAGTGCTGGAAACATTTGACAA	2735
Qy		905	AlaCysArgAsnLeuLysProLeuGluValLysProHisGluThrValThrcGluSerThr	924
Db		2736	ACATGTCACGATCTGAAACCTCTGGAGGTGAAGTCGCATGAAATATGTACAGAAAGCAC	2795
Qy		925	ThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThrGluCysLysSer	944
Db		2796	ACGACGACCAACAACGACAACAACGACCGTTTACCGATCCGAAGGCAACGAATGCAAATCC	2855
Qy		945	LeuGlnThrThrAspThrThrTryValThrcGlnThrSerThrHisThrSerThrSerThrIle	964
Db		2856	TTACGACACAAACGATACATGGGTTTACACAGACATCGACACACACACGACCTTACCTC	2915
Qy		965	ThrSerThrIleMetSerLysIleThrLeuThrSerThrArgArgCysLysProThrLys	984
Db		2916	ACATCTACATCACTCAAAAATATACATTGACATTAACGAGCGCATGCAACCAACCAAG	2975
Qy		985	CysThrThrGlyGluGluAspAspAlaGlyAspValLysProSerGluGluLeuArgMet	1004
Db		2976	TGTACGACAGGG-----GATGAACGACAGAGACGTGAACCGAGTGGAAGATG	3029
Qy		1005	SerGlyTPanValMetArgGlyValIleValAlaMetValIleSerPheMetIle	1023
Db		3030	AGTGGGTGAGCGTGATGAGGGGGGTGATAGTACGAAGTGTATTTCCGTTATCATGATT	3086

RESULT 9

PMCSGI	PMCSGI	3363 bp	DNA	linear	PLN 26-SEP-1994
DEFINITION	Pneumocystis carinii B-cell receptor (meg1) gene, 3' end.				
LOCUS	L27092				
ACCESSION	L27092.1 GI:535706				
VERSION	B-cell receptor.				
KEYWORDS	Pneumocystis carinii				
SOURCE	Pneumocystis carinii				
ORGANISM	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;				
REFERENCE	Pneumocystidaceae; Pneumocystis.				
AUTHORS	1 (bases 1 to 3363)				
TITLE	Garde,T.R. and Stringer,J.R.				
JOURNAL	Molecular characterization of clustered variants of genes encoding				
MEDLINE	major surface antigens of human Pneumocystis carinii				
PUBMED	Infect. Immun. 62 (8), 3092-3101 (1994)				
COMMENT	94314421				
FEATURES	7518806				
SOURCE	Original source text: Pneumocystis carinii DNA.				
	Location/Qualifiers				
	1..3363				

gene	CD5
<pre> /organism="Pneumocystis carinii" /mol_type="genomic DNA" /db_xref="taxon:4754" 152..3244 /feature="mggi" &lt;152..3244 /feature="mggi" /codon_start=1 /product="B-cell receptor" /protein_id="AAA21645.1" /db_xref="GI:535707" /translation="VARAAKRVAGVKNNAEERLPALIMRADYKDESKCKNKI DGLKNAISLSEEVHKEIDPCDKSGGCKGKELKKNVAKKNNKFTLEGLVKKDAS LINDCKENREKPCLEBACAPDLYEDSKLNLCTYCKRBSVAEVLRYLRRLDGN TECEKKIKDVCPKIQESDELTLCLDCKCTNLMPARDKCNLTLEEDVKKLEEN NLGKCLPLLEHATPTEGTAKKASQCTPNKDCYLRPCDELABECSKGI1YIHPK DEPTEPPEVLEADIGLELYKKAEDVHIGKPPVADATYLAAL1ONDPKIOANA KECKCKVLEKCKELKKEHVEGLDLCNONAASOGSKCELEKEKASTKILSEKI KILSKSGERTIPMYKLTSTPLSDPCARLESDCFYPAQKDP1KKECKNVAKCYKGI AARAKVQENMGRLRGNSQSLKKFQOELKYVCEKLEBKESFSFDELPLVCQD KAARLTGDLRMVTL1FLRQOLDQRDPFTVTAAG1KEKCDLKGFKQK1TWCHQ BOCKRLGDTLETLKOVLLNENKDTLKTENCVTL1KCKNMKSRGDRSPVCFE ATCKLMMVQVORCK1FENIKVSEI1VDF1KNNNN1ITLERNPSMAY1CNFSFS POPSKNPCTK1KNNCKP1FYERKLEDA1KYLEKGLSDEKCTAAL1KQCTLAAG NMSVSLCKDNTOGSKNKTDENKVVLECKLMEKVKOCERTPL1EKLORDP1LEK TYBELKANKKAMKNSLVLSVFKVDGNPDKNNKSKEDKVVNSKOT1KHKYILK GVNDVLTLEAKKAPDLAAEFGVYD1KEKCEK1LDCQ1KDDCDEKVCQK1IKK CNDLKLKELSHSEI1VTESTTTTTTTTTTTTTVTD1PKATCKSLQTD1TWTOT1STH1ST INSTISK1TLT1STRCKP1KCTTGDAD1EDVK1PSEGLRVSGMNVIRGAI1VAW1ISFW " 152..3241 /feature="mggi" /product="B-cell receptor" /standard_name="major surface antigen" /evidence=experimental </pre>	<pre> /organism="Pneumocystis carinii" /mol_type="genomic DNA" /db_xref="taxon:4754" 152..3244 /feature="mggi" &lt;152..3244 /feature="mggi" /codon_start=1 /product="B-cell receptor" /protein_id="AAA21645.1" /db_xref="GI:535707" /translation="VARAAKRVAGVKNNAEERLPALIMRADYKDESKCKNKI DGLKNAISLSEEVHKEIDPCDKSGGCKGKELKKNVAKKNNKFTLEGLVKKDAS LINDCKENREKPCLEBACAPDLYEDSKLNLCTYCKRBSVAEVLRYLRRLDGN TECEKKIKDVCPKIQESDELTLCLDCKCTNLMPARDKCNLTLEEDVKKLEEN NLGKCLPLLEHATPTEGTAKKASQCTPNKDCYLRPCDELABECSKGI1YIHPK DEPTEPPEVLEADIGLELYKKAEDVHIGKPPVADATYLAAL1ONDPKIOANA KECKCKVLEKCKELKKEHVEGLDLCNONAASOGSKCELEKEKASTKILSEKI KILSKSGERTIPMYKLTSTPLSDPCARLESDCFYPAQKDP1KKECKNVAKCYKGI AARAKVQENMGRLRGNSQSLKKFQOELKYVCEKLEBKESFSFDELPLVCQD KAARLTGDLRMVTL1FLRQOLDQRDPFTVTAAG1KEKCDLKGFKQK1TWCHQ BOCKRLGDTLETLKOVLLNENKDTLKTENCVTL1KCKNMKSRGDRSPVCFE ATCKLMMVQVORCK1FENIKVSEI1VDF1KNNNN1ITLERNPSMAY1CNFSFS POPSKNPCTK1KNNCKP1FYERKLEDA1KYLEKGLSDEKCTAAL1KQCTLAAG NMSVSLCKDNTOGSKNKTDENKVVLECKLMEKVKOCERTPL1EKLORDP1LEK TYBELKANKKAMKNSLVLSVFKVDGNPDKNNKSKEDKVVNSKOT1KHKYILK GVNDVLTLEAKKAPDLAAEFGVYD1KEKCEK1LDCQ1KDDCDEKVCQK1IKK CNDLKLKELSHSEI1VTESTTTTTTTTTTTTTVTD1PKATCKSLQTD1TWTOT1STH1ST INSTISK1TLT1STRCKP1KCTTGDAD1EDVK1PSEGLRVSGMNVIRGAI1VAW1ISFW " 152..3241 /feature="mggi" /product="B-cell receptor" /standard_name="major surface antigen" /evidence=experimental </pre>



[illegible]

Dh	61	GAAGAAATCTTTGGCTTTAATTGACGAGAAAAAATATATATGATATGAATCAAAAAA	120
Qy	39	LyLeuLyLeuLySfYrCYsGInGuLeuThrGIuAlaLyLeuLeuLeuIleGIuGIuValHis	58
Dh	121	GAATTGAAAAAATATTTGTAAGACATTAAACGAGATGCAAAATTTAAACCGAAAAAGTTT	180
Qy	59	ArgLyLeuLeuLySglPheCYsGInuAAspGIuLyAlaAspThrLySGLySGLyLeuLyS	78
Dh	181	AAAAAACTTAAAGAGTTTGTGAAAAATAAAAAGCAGATTCAAAAATGCAGAAAGATCTGAA	240
Qy	79	AlaAsnIleGIuLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	98
Dh	241	GA AAAAATCTCACTCAAAAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA	300
Qy	99	LySGLIleGIuIleIleThrAspLySAspCYsLySGLySGLySGLySGLySGLySGLyS	118
Dh	301	AAAAATTCAGATTAAACCGATGGAAGATTCGAAAGAGATGAAACAAATCTCTATTTTGG	360
Qy	119	GIuGIuLyAlCySserLySGLyLeuLySAspAspCYsAsnThrLeuArgLeuLySGLyS	138
Dh	361	GAGGAGACATGCTCCAGCGCAACTTAAAGATGATTGCAATCTTTGAGAAAATTAAGTGTAT	420
Qy	139	GIuLySGLySAspLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLy	158
Dh	421	CAAAAGAAAGCTGATTAAGTGGCGGAAGAGCTCTTTAAGACACATCTCTGAGAGCTTA	480
Qy	159	AsnGIySerValIleCYsGIuLySGLyLeuLySGLyIleCYsAspValMetGIyArgGIu	178
Dh	481	ATCAATAAACTACATGTCATGGAAGGAAAGCTCAAGAGGTTTGATGAGTTGAGTCAAGAA	540
Qy	179	SerAspGIuLeuThrAsnLeuCYsLeuAsnGIuLySGLyThrCYsLySAsnIleLeuIle	198
Dh	541	AGTATAGATTACGAAAGCTTTGTCTTTATCAAAAAATGACGTGCAAAACATTTGTATTA	600
Qy	199	GIuLySAspLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	218
Dh	601	GA AAAAACA AAAAATTAATGCTCTTTAAACAGAGTTTAACGACAGACTTGAC-----	654
Qy	219	LySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	234
Dh	655	AAGAAAGATGAGTTACGAGGAAATGTTTACCACTGCTGAACGATCTATTTTATAGA	714
Qy	235	GIyAsnCYsGLySAspAspAspIleIleLySGLySGLySGLySGLySGLySGLySGLy	247
Dh	715	GGGAAATTGT-----GAAGATATATCAAAATGTAAATCAATCCGAGACTGTATGAA	768
Qy	248	-----LeuGIuLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	260
Dh	769	TATTTGCCAGTGTGTATACATTGGCAGTGAATGTGAAGAAATTAAGATTATTATACA	828
Qy	261	ProGIuLySAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp	280
Dh	829	CATCGGGAGATCCGATTTCAATCCAACTAAAGTCAAACTGTAAGCAAGACATAGGA	888
Qy	281	LeuGIuGIuPheLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	300
Dh	889	CTGGAAGAAGCTTTATATAAAAGCGCGCAAGAAAGTGTTCATATGTGAAAAGCCTCTGTA	948
Qy	301	ArgAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp	320
Dh	949	AGAAATCACTGCTCTACTGGCGCTTTATATCAAAATCTAGATCTTAAGAGTCAAGTCTG	1008
Qy	321	AspLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	340
Dh	1009	GGTAAAGAA---TCCGAAAAAGTTCTTAAGATTAAGTAAAGATTAAGATTAAGATTAAG	1065
Qy	341	AlaLeuGIuSeriLeuCYsLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	360
Dh	1066	ATTTTGGAGATTTTGTATCAAAAATGTAGCTGTCTCAAAATGAAATTTGAAAAAGTGTAA	1125
Qy	361	GIuLeuGIuAsnAspIleAsnLySGLySGLySGLySGLySGLySGLySGLySGLySGLyS	380
Dh	1126	GAGTTTAAGAAAGAGTTAGCAAAACAGTAAATTTTGTGAAAAAATTAAGAAATTAAG	1185



QY 381 ArgLeuPheAerProThrLyGlyAsnAangLileValGlyTrpGluGlyLeuProThr 400  
 DB 1186 CACCTC-----TCGGATCCGAGAGACTCATTCATGATTAAGTACACACA 1233  
 QY 401 PheLeuSerAangLyuAerCyValAluLeuGluSerTyCyPheTyPheGluLyLe 420  
 DB 1234 TTTCTTAGTGCATGACTGCACAGGTTAGTACAGCTGTTTTTTATTTAAAAAGTCAA 1293  
 QY 421 CyPro---AerGlyGluAenAlaCyLeuAenAlaLeuGlaAthrCyGlyTyLeuAerGly 439  
 DB 1294 GCACCTCTTGCACAAAGAA-----TGTAAATCTGAAGGACGATGTTAAAGAGGG 1347  
 QY 440 LeuAerAlaAerGlaAenLyValLeuGluGluAenMetAerGlyMetLeuHieGlySer 459  
 DB 1348 CTTCAGACACAGGATTAAGACATTCGCAAAAAAGATGTCAGACTGTTTATGCTTCA 1407  
 QY 460 AenLySerTrpLeuGluLyPheGluGluGluLeuValCyGluLyLeuLy 479  
 DB 1408 GGCAGAAAGATGTTTAAAGAACTAGTAAATAATGAAGAAATGTCGAACTTAA 1467  
 QY 480 LyGluGluAenLyGlySerPheSerAenAerGluLeuPheLeuLeuCyValGlnProAla 499  
 DB 1468 ACA-----ACAAAGCATGAGTGTGTTTGTCTATGATTAATGATCCACTT 1509  
 QY 500 LyValAlaAerGluLeuThrHieAerPheAerMetLySerThrIlePheLeuAerGluGln 519  
 DB 1510 AAACAGCTCAGAAATACCTTCGAGTGTATTCACAGCAAGACAGCTTTTTCGGAACAA 1569  
 QY 520 LeuAerGluLyAerAerPheProThrAerLyAenCyLyGluLeuGlyAerGlyCy 539  
 DB 1570 TTGATCAAAAGCAGACTTCCAAAGACAGAAAGATTGCAAGAAATTAAGAAAGAGT 1629  
 QY 540 GlnAerPheGluGlyAerSerLyGluIleThrTrpProCysHieSerLeuGluGln 559  
 DB 1630 GAACCTTTAGGGAAGATTCAAAATCAGATTAAAGTGGCCATGTCATACCTTAAACACAG 1689  
 QY 560 CysAenAerGluGlyThrGluIleLeuLyGluValLeuLeuAerGluHieLyAerP 579  
 DB 1690 TGTATGCTTTGGGACTACAGAAATCTTGAAACAGGTTTACTAGATGAACACAGAT 1749  
 QY 580 ThrLeuLyAerGluGluSerCysValLyTyLeuLyGluLyCysAenLySer 599  
 DB 1750 ACTTTAAGAACTAGTAAACTGTACGAATATTAAAGAAATATGCATTAATGGCT 1809  
 QY 600 ArgArgLyAerAerAerPheSerPheValCyValPheGluAenAlaThrCyGluLeu 619  
 DB 1810 AGAAGGGGTGATGATCGTTCTTTGTATGTGTACCAAAACGCTACGTGAAGCTG 1869  
 QY 620 MetValLyAerPheValLyAerAerCyGluValPheLyAenAlaLeuAlaSerTy 639  
 DB 1870 ATATGATGATGTGTGAAGACAGGTGAATTTGAAAAAATATGCCAACGCTCAGAT 1929  
 QY 640 IleIleGluPheLeuGluAenAenThrAenLyIleThrThrLeuGluAerAerCyPro 659  
 DB 1930 ATTAAATATCTCTTAAATTAATTAACAAATTAATTAAGTAAATTAATGATTCGCA 1989  
 QY 660 SerTrpHieThrTyCyAenAerPheSerProAenCyProGlyLeuThrLy---Glu 678  
 DB 1990 TCATGACACCCATCTGCGATAGATTTTACCCTGATCTTAAAGAAAGGAA 2049  
 QY 679 AenSerCyThrLyValIleValLyHieAerGluProPheTyLyAerGluValAlaLeuGlu 698  
 DB 2050 ACTTTCTGTCAAAATCTTAAATAATATGCGAACCATTTCTCAAAAGAGGATTTTGA 2109  
 QY 699 AenAlaLeuLyValAlaLeuGluGluGlyLeuThrAerLySerTyCyGluProAla 718  
 DB 2110 GATGCTCTTAATAGTACGCTTCAAGGGAATTAAGTAAATTAATTAATGATTCGCA 2169  
 QY 719 LeuLyAerGlyTyCyThrValAlaGlyAenValAenAenAlaSerIleSerGlyLeuCy 738  
 DB 2170 TTAAAGAAATATTCACAAATATTTGAAAAATGAAGTATCATCATCAACAGTTTATGT 2229

QY 739 LyValAenThrLyAerAenSerGlyLySerAerGluAerAlaAerGlyGluLeuCy 758  
 DB 2230 AAAGATAAATACCGAAGTAATAAACTAAAAAGCCGATTAATGAAGTTAGAAAGACTTTGT 2289  
 QY 759 GlnLyLeuValLyGluValAlaGluGluGlnCyValAlaLeuProThrGluLeuGlyGln 778  
 DB 2290 CTAAATTAATGTGGAAGGAGGGAACAGCAATGTAAATGTTCACGACGAATGGACAT 2349  
 QY 779 ProAlaAlaAerPheLyAerTyLeuThrTyGluGluLeuLyAerGluAlaGly 798  
 DB 2350 GAGCAAAAAAGCTTAAAGATGATTTTGAAACATTTGAAAACTTAAAAACAGGACGAG 2409  
 QY 799 GluAlaAerAenLySerSerLeuValLeuSerLeuIleLyLyAenGluSerAerVal 818  
 DB 2410 AAAACATGATTAATATCAATCTTTTATCATCTTAAAGAAAGATGAATAATATACA 2469  
 QY 819 SerLySerAenSerLyAenLyAerLyAerValAerAlaValSerAenGlyLeuGlnAerThr 838  
 DB 2470 TCGAAAAATGTAGCAAAAGCAAGGATTAAGATACCTTCAACGACCTTCAAGATACC 2529  
 QY 839 ThrLyHieValLyIleLeuAerGlyValLyAerValSerValThrGluLeuGlu 858  
 DB 2530 ACAGAACTATGAATAATACATACGAGAGAGATTAAAGATGTATCCATACGAATTCGAA 2589  
 QY 859 AlaLyAlaAerPheAerPheAlaAlaGluValPheGlyAerGlyValAerPheLyGluAer 878  
 DB 2590 GCTAAGGCAATTAATGATTTGATAGCAGAAATTTGAAGATATCTAAGCTTGAAGAAAGA 2649  
 QY 879 CysAenLyLeuGluGluSerAerCyAerGlyLeuGluAerCyLyAerPheLeuGluVal 898  
 DB 2650 TGTAAATAATGTGATCAGATTGCGAGATTAAAGGAGATTGCAAGGATTTAAGAAAGAGTA 2709  
 QY 899 CyLyLyLyIleAenLyAlaCyAerAenLyAerProLeuGluValLyProHieGlu 918  
 DB 2710 TGTGAAAAATACAGAGATGATGTTGAAATTTAAACCATGAAAGTAAAGCCGACAGAA 2769  
 QY 919 ThrValThrGluSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 938  
 DB 2770 ACGGTGACAGAAAGCACAAGACGACCAAGCAACGACCAACGACCTTACGATCCGAA 2829  
 QY 939 AlaThrGluCyLySerLeuGluThrThrAerThrTrpValThrGluThrSerThrHie 958  
 DB 2830 GCAACAGAAATCAAAATCTTTACAGACAAAGATTCATGATTTACACAGACTTCACACAT 2889  
 QY 959 ThrSerThrSerThrIleThrSerThrIleThrSerTyIleThrLeuThrSerThrArg 978  
 DB 2890 ACCGACAGCTTACATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2949  
 QY 979 ArgCyLySerProThrLyLyCyThrThrGlyGluGluAerAerAlaGlyAerValLyPro 998  
 DB 2950 CGTTGCAAAACCAACCAAGTGTACAGCGG-----GATGATGACAGAGAGCTGAACCG 3003  
 QY 999 SerGluGlyLeuAerMetSerGlyTrpAerValMetArgGlyValIleValAlaMetVal 1018  
 DB 3004 AGTAGGAGATTGAAGATGAGTGGGTGAACGATGATGAGGGGGGTGATAGTAGCAATGTT 3063  
 QY 1019 IleSerPheMetIle 1023  
 DB 3064 ATTTGTTTCATGATT 3078  
 RESULT 11  
 AF033212 3080 bp DNA linear PLN 10-SEP-1998  
 LOCUS Pnuemocystis carinii f. sp. hominis clone HUNSG32 major surface  
 DEFINITION glycoprotein (MSG) gene, partial cds.  
 ACCESSION AF033212  
 VERSION AF033212.1 GI:3560520  
 KEYWORDS Pnuemocystis jirovecii  
 SOURCE Pnuemocystis jirovecii  
 ORGANISM Pnuemocystis jirovecii  
 BUKARYOTA; Fungi; Ascomycota; Pnuemocystidomycetes;  
 Pnuemocystidaceae; Pnuemocystis.  
 REFERENCE 1 (bases 1 to 3080)

AUTHORS Garbe,T.R. and Stringer,J.R.  
TITLE Molecular characterization of clustered variants of genes encoding  
JOURNAL major surface antigens of human Pneumocystis carinii  
MEDLINE Infect. Immun. 62 (8), 3092-3101 (1994)  
PUBMED 94314421  
2518806  
REFERENCE 2 (bases 1 to 3080)  
AUTHORS Mei,Q., Turner,R.E., Sorial,V., Kilmington,D., Angus,C.W. and  
Kovacs,J.A.  
TITLE Characterization of major surface glycoprotein genes of human  
Pneumocystis carinii and high-level expression of a conserved  
region  
JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)  
MEDLINE 98380374  
PUBMED 9712777  
REFERENCE 3 (bases 1 to 3080)  
AUTHORS Mei,Q., Turner,R., Sorial,V., Kilmington,D., Angus,C.W. and  
Kovacs,J.A.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National  
Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD  
20892-1662, USA  
FEATURES  
source Location/Qualifiers  
1..3080  
/organism="Pneumocystis jirovecii"  
/mol\_type="Genomic DNA"  
/specific\_host="Homo sapiens"  
/sub\_species="hominis"  
/db\_xref="taxon:42068"  
/clone="HUMSG32"  
/note="derived from HIV-infected human with P. carinii  
pneumonia"  
<1..3029  
/gene="MSG"  
<1..3029  
/gene="MSG"  
/note="surface antigen"  
/codon\_start=3  
/product="major surface glycoprotein"  
/protein\_id="AAC34975.1"  
/db\_xref="GI:3560521"  
/translation="ARAYKROAKRSGSVYDGEIILALINAGKKNNECKEKELEK  
CKTLIDALEKPEKVKLKEFENKKAUSCKELKEKLTQKCTATKGLTAIKKNS  
DLTDEDCKENQOCLFLBACPAELKDCCNTLRNKCYQKRDKVAEALIAVAGLI  
NETTCBGLKEVCIELSQSDDELTYLCYQKCTFLYLEKQKCNALAKOVNALAEK  
KDELRGKCLPLERCFYRGNCEIDISKNSSEDCYFLPCVDTLAVCEENKIYTH  
PGSDNPRTSKSPYVAEDIGLBELYKKAEBGVHIGKPRVATLALILIONLPKSO  
VGKEERKLNCKEKLKSHETLIDPCNONVAGNIECKELKEKLNASTITLPEKIK  
NHLISGSEVLPMTLTLPFLSDNDCTRLSDCFYKSGAPLDKECNLKAACVKGLE  
AQNALQKKNYGLFYVSGKEMFKLLKIMECESELTTSDBELFLCIDPLKAVRL  
AADIQARAVFLRKQIDKRPDPTDKCELRKCEALGKDSNOIPKCHTKOCDRL  
GTTTELKQVLDENKODLTATHENCTYLRKCHKHRSRGDRSFVQYONATGLIV  
DDVDRGCVPEKNQASDINSLKMKOIKTSSAANICPSMHPYCDRFLPNDLKKG  
TFCQNLKRYCEPFYRKLTEDALKVELQNSNRKCSALERTYITLIKNSDSINS  
LCKDNTSKTKTDNEVRKLCCLKVEVEEQCKMLPDLHEEDKLDDEYTEKIK  
KQAEKTNKSNLVLSFVKDEENYTSKNSKDKKNVANGDQDTEHAKILIRGKDV  
SLTSEBAKAFDLVAEVPGRYLDLKERCKNLSDCKVKNDCQDLBEGVCGKIGVSKLK  
PLKVPHEVTESITTTTTTTTTVTDPRATECKSLQETDWTITQSTHSTSTSTST  
TSKTLTSTRCKPTKCTTGDAEDVKRSEGLKMSG"  
3029  
/gene="MSG"  
/note="putative PCR artifact leads to premature stop  
codon"

ORIGIN  
Alignment Scores:  
Pred. No.: 9,68e-210 Length: 3080  
Score: 3396.50 Matches: 654  
Percent Similarity: 76.15% Conservative: 141  
Best Local Similarity: 62.64% Mismatches: 208  
Query Match: 62.77% Indels: 41  
DB: 8 Gaps: 12

US-10-654-416-14 (1-1023) x AF033212 (1-3080)  
QY 2 ALaarGlaVallysaTarglnAlaIaglyThrGlnaSer-----IleaspGluGlu 19  
Db 3 GCGGGCGGCTCAAGCCGCGAGCTGCAAAAGCATCAGACGCTAGTATGATGCTGAA 62  
QY 20 HisValleuAlaleuIleleuLyseGluaspGlyleuSerGluGluGlyLyse 39  
Db 63 GAATTCCTTTGGCTTAAATGTCAGGAAAAAATATATATGATGATGAAAAAGAA 122  
QY 40 LeuLyseLyseTyrcyGslngluLeuThrgluAlaLyseleuAenIleGluGlnValHisary 59  
Db 123 TTGAAAAAATTTTAAAGACATTAAACGATGACAAATTAAAAACGAAAAAGTTCA 182  
QY 60 LyseleuysgLyPheCyseGluaspGlyLyseAlaapThryscyLyseGluLeuysAla 79  
Db 183 AACTTAAAGAGCTTTGGAAAAATAAAAAGCAATTCAAAATGCAAGAACTGAAAGA 242  
QY 80 AenIleGluLyseLyseCethThrlleLyseGlyLyseleuysgLuAlleLyseLyse 99  
Db 243 AACTCACTCAAAAATGTACTGCAATCAAGGAAACCTTACAGAAAGCAATCAAAAAA 302  
QY 100 IlleGlnIleIethrAapLyasapCyseLyseGluasngluGlnGlyPheleuGlu 119  
Db 303 AATTCAGATTTAACGATGAAGATTGCAAGAGAAATGAACAATGCTCTTTTGGAG 362  
QY 120 GlyValCyseSerLyseGluLeuLyasapCyseAanthrleuArgaenLyseTyrgln 139  
Db 363 GGACACATGTCAGCGGAACCTTAAAGATGATGCAATCTTGAAGAAAAATGAGCTCA 422  
QY 140 LyseLyseAapLyseValAlaGluGluValleuLeuArgAlaLeuAgsaeAapLeuA 159  
Db 423 AAGAAAGGTGATTAAGGCGGGAAGAGCTTTTAAAGAGCAAGTTCGAGGCTATC 482  
QY 160 GlyseValIleCyseGluLyseLyseleuIleCyseProValMetGlyAysGlyser 179  
Db 483 AATGAACATCATGTGAAGCAAACTCAAGAGGTTTGCAATGAGTTGATGCAAGAAAGT 542  
QY 180 AspGluLeuThraenLeuCyseLeuasnGlnLyseGluThrcyLyseAenIleleuIleGlu 199  
Db 543 GATAGATTAAAGAGCTTTGTCTTATCAAAAAATGACGTCAAAAACCTTTGTTAAGA 602  
QY 200 LyseAapLyseLyseGlyThrleuLyethraapValserAlaIaleuGlySerPheLy 219  
Db 603 AAAAATAAATAATGATGCTTTAAACAGATGTTAAACGACGACTTGAG-----AAG 656  
QY 220 LyseGlu-----ThrcyaleuGluLeuGluGlnGlyTyrcyPheTyrlleGly 235  
Db 657 AAGATGAGTTACAGAGAAATGTTTACCACTGCTTGAACGATGCTATTTTATGAGGG 716  
QY 236 AenCyseGlyAapAapAenIleIleLyseCyseIleGlu----- 247  
Db 717 AATTGT-----GAAGATATATCAAAATGTAAATATCATCGAAGACTGTATGATAT 770  
QY 248 -----LeuGlyGlyLyseCyseGlnGluAenIleAlaTyrcPro 261  
Db 771 TTGCGAGTGTGTATACATTGCGAGTGAATGTAAGAAATAAGATTATTAATACAT 830  
QY 262 ProGlyProAapPheapProThraPProGluAlaThrlleAlaGluAapIleGlyLeu 281  
Db 831 CCGGATCCGATTCATCAATCAATGATGCAAGGCTTACTGTAGCAAGAGACATGAGACTG 890  
QY 282 GluGluPheTyrllyLyseValaGluGluAapGlyValPheIleGlyLyseAenIleleuA 301  
Db 891 GAAGAGCTTTAATAAAGCCGCGAGAGAGAGTGTATATGGAAGCCCTCTTAAGA 950  
QY 302 AapAlaThraAlaLeuAlaLeuLeuIleGlnAapSerSerLeuLyseLyseAapAap 321  
Db 951 GATGAACTGCTTACGCGCTTTGATTCATAAATCTAGATCTCTAAGAGCAAGTGGGT 1010  
QY 322 LyseGluLyseCyseGluGluAlaLeuGlnLyseCyseLyseAapProHisleGluHisleGluAla 341  
Db 1011 AAAGAA--TGCAGAAAGATTCTTAAGATATAGCTTAAGAGTTAAAAAGTCAAGAAATT 1067

Oy	342	LeuGIuSerLeuCyVbLyLeVaAnGIuLeuSerAenAerGIuThrLyLeCyVgGIuGIu	361
Db	1068	TTGGGAGATTTTTGTATCAAAATATGACTGCTCAAAATGAAATTGAAAGGTGAAGAG	1127
Oy	362	LeuGIuAenAerPILeAenLybThrCyVbLyIlePheThrSerLyVaIleThrAenAer	381
Db	1128	TTAAGAAAGAGCTTAAGCAACAGTACTAAATCTTTTGAAGAAAATTAAGATTAACAC	1187
Oy	382	LeuPheAerProThrLyLeGIyAenAenGIuIleValGIyTPRGIuGIuLeuProThrPhe	401
Db	1188	CTC-----TCTGGATCCGGAGAGCATTCATCGATGATTAAGTTAGCGACATTT	1235
Oy	402	LeuSerAenGIuAerCyVaIleLybLeuGIuSerTyCyVbPheTyRPhGIuLybCyVg	421
Db	1236	CTTAGTGCAATGACTGCACAAAGTTTAAGTCAACTGTATTTTATTAATAAGTCAACA	1295
Oy	422	Pro--AerGIuGIuAenAIIeCyVbLyAenIleAerGIuIleAerThrCyTyLyAerGIuLeu	440
Db	1296	CCTCTTGCAAAAGAA-----TGTAATAATCTGAAGCGACATGTTATTAAGAGAGCGTT	1349
Oy	441	AerPIleAerGIIeAenLybValIleuGIuIleuAenAerGIuMetLeuIIeGIySerAen	460
Db	1350	GAAACACAAAGCTTAATGAAGACATTCGAAAAAAGATGACGACATGTTATGTTCAAGC	1409
Oy	461	LybSerTPRLeuGIuLybPheGIuIleuLeuValLybValCyVbGIuLybLeuLybLy	480
Db	1410	AAAGAAATGGTTTAATAAACTACTAATAAAAAATTAAGAAAGATGTTCCGAACTTAACAA	1465
Oy	481	GIuAenLybGIySerPheSerAenAerGIuLeuPheIleuCyVaIGIuProAlaLy	500
Db	1470	-----ACAAGCATAGTTGTTTGTTCGTAATGATTAATGATTCACCTTAAA	1511
Oy	501	AlaAlaAerLeuLeuThrIIeAerLeuAerGlyeThrIIePheLeuAerGIuIleu	520
Db	1512	GCAAGCAAAATACCTTGACAGCTGATATCCAAAGACAGACAGCTTTTTCGGAAACAATTG	1571
Oy	521	AerGIuLybAerPheProThrAerLybAenCyVbLybGIuLeuGIyAerGIySerGIu	540
Db	1572	GATCAAAAGCGAGACTTTTCCAAACAAAGATTTCCAGAAATTAAGAAAGAAAGTGTCA	1631
Oy	541	AerLeuGIyGIuAerPheSerLeuGIuIleThrTPRProCyVbIIeThrLeuGIuIleuCy	560
Db	1632	GCTTATGGGAAGGATTCAAATCAAGTTTAAGTGGCAATGTCATACGCTTAAACACACAGCT	1691
Oy	561	AenAerLeuGIyThrThrGIuIleuLybGIuValLeuLeuAerGIuIIeLybAerThr	580
Db	1692	GATGCGCTTGGGGACATACAGAAATCTTGAAACAGGTTTAACTAGATGAACACAAAGATCT	1751
Oy	581	LeuLybAerGIuIleuSerCyVaIleLybTyRLeuLybGIuLybCyVbAenLybTPSerAer	600
Db	1752	TTAAGAACTCATGAAACTGTATCGAAATATTTAAAGAAAAATGCTCAAAATGCTGTGA	1811
Oy	601	AerGIyAerAerAerPheSerPheValCyVaIlePheGIuAenAIIeThrCyVGIuIleuMet	620
Db	1812	AGGGGTGATGATCGTTCTCTTTGTATGTGTTTACAAACAGCTACGCTAAGCTGTATA	1871
Oy	621	ValLybAerValLybAerAerCyVgGIuValPheLybAenIleAerIleAerTyRILE	640
Db	1872	GTAAGTGAATGTGAAGACAGCGTGTGAAGTATTGAAAAAAATATGCAAGCGTCAATATT	1931
Oy	641	IleGIuPheLeuGIuAenAenThrAenLybIleThrThrLeuGIuAerAenCybProSer	660
Db	1932	AATAATCTCTTAAATAATTAACAAATTAACAAACAGATACGACGAAATATTTGCTCCCA	1991
Oy	661	TPRIIeThrTyRcyVbAenAerPheSerProAenCybProGIuLeuThrLyb--GIuAen	679
Db	1992	TGGCAACCAATATGCGATAGATTTTACCAATATGCTCATTTAAAGAAAGAAAACCT	2051
Oy	680	SerCybThrLybIleLybLybAerGIuProPheTyRLebAerLybAlaIleuGIuAer	699
Db	2052	TTCTGTCAAAATCTTAAATAATTAATTCGAAACCATTCCTAACAAAGAAAGTTTAAAGAT	2111

[illegible]

VERSION	AR438434.1	GI:42663289
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 3042)	
AUTHORS	Kovacs,J.A., Huang,S., Masur,H., Fischer,S.H., Gill,V.J. and Mel,Q	
TITLE	Identification of a region of the major surface glycoprotein (MSP)	
JOURNAL	gene of human Pneumocystis carinii	
FEATURES	Patent: US 6664053-A 1 16-DEC-2003,	
SOURCE	Location/Qualifiers	
	1..3042	
	/organism="unknown"	
	/mol_type="genomic DNA"	
ORIGIN		
Alignment Scores:		
Pred. No.:	7,476-169	Length: 3042
Score:	2762.50	Matches: 541
Percent Similarity:	69.10%	Conservative: 177
Best Local Similarity:	52.07%	Mismatches: 280
Query Match:	51.05%	Indels: 41
DB:	6	Gaps: 18
US-10-654-416-14 (1-1023) x AR438434 (1-3042)		
OY	1 Metalaargalavalaylsyarginalaalaglythrlnasnerlleapnglunhs	20
Dd	1 GTGGCGCGGCGGTTAAGCGCAGGTAAACAGAGCA--TCAGGAATGATGAGAGAA	57
OY	21 Val-----Leualauleulleulsylglaabpgilyleuserglunglucysyllys	38
Dd	58 GTGCCTCTTTTGCTTAACTAATAAAGAAGATTCTAGAGATGATAAAAATGCCAAGAA	117
OY	39 LysleuluylsyetrCysegsluneuthrgualalyleuasnllegluglnvalhs	58
Dd	118 AATTAAGAAAACATTGCACAAGATTAAAGGAAGCAACTTAATCCAGAACAAAGTACT	177
OY	59 ArglyLseuluylsyGLYPheCysglunabpGilylsyalaaSpThrlyscysylsgluelys	78
Dd	178 GAAAAGTTAAAGATTTCGTGTATGACAAAACCGATTAATAAATGTAAGAATCTAAA	237
OY	79 AlaenllleglluylsyCsyrThrIlelysgilylseuuleuglualelylys	98
Dd	238 AAAATGTTGAAAAAAAATGCGGTGATTTTAAACAGAAATTGAMAAATTGTGAAAAAG	297
OY	99 LysllleglnllelthrapslysaapCybsylsglaanglnglnglnCysleupeleu	118
Dd	298 GAAGCTTCAAATTTGAAAAATGATGATGATCAAAAAATGAACACAGTGTCTTTTAA	357
OY	119 GluglyValCyseSerlysglupeulysaaPspCyasaThrleuargAsnlyscsTyR	138
Dd	358 GAAAGAGCATGCTCT---GATCTTCAAAAGAAATTGCAAGATTTAAGAAACAAATGTAT	414
OY	139 GlnlylsyarsApIysValaaaglunluvalleuileuarglaaleuarg-----	155
Dd	415 CAGATTAAGCCGTGATAGGTACCAAGGAAGTCTTTTAAGAATATATAAAGGAAGAAT	474
OY	156 ---SerapleuanenglyserValleCysegilyulylyleuuleugluileCysProval	174
Dd	475 TTTAAAGATAAAAATTTCA-----TGTGAATAATAACTGAAATATACGTCAAGAA	525
OY	175 MetGlyArgGlusefarpGluethraenleuCyleleuansnglnlysglutHrCyslys	194
Dd	526 TTAAAGTCAAAATGATGACGAAATTGAATGAATTTATGTTTGAATCAAAAAATACGTGTAT	585
OY	195 leuileleullelgilylsaplylsyCsylglyThr-----leuylsrhaepVal	211
Dd	586 AATCTTGTAAAGAAACGACAAAGGTGGAATCTTTCAAAAATCTTTAAAAAGGAAT	645
OY	212 Serallaalleuglyserhelys---lysglutHrCysleugluileuuleuglnglnCys	230
Dd	646 AAACAAATTAAGAGATGAACAACATTAATAAAAAATATGCCATTAATTAATGAAGATGC	705

QY	231	TyrPheTyrTlleGlyAenCyvGgLYaBpAaBpRl1e1LeuScyS1leGluLeuGlyGly	250
Db	706	ATTGTTTATGATGAAAGTTGTGGAAAGAAATTC--CTGAAGGTAGTGAATTGGAAAA	762
QY	251	LYvCyGngInGInaEni1eA1aYMetProProG1yProAaBpPheAaBpProthArg	270
Db	763	AAATGTCAAGAGAAATATATTACTTACACTTATCATATTCAAGGTTTGATTCCTATGAA	822
QY	271	ProG1uA1aThr1leA1aG1uAaBp1leG1yLeuG1uBpHeTyrLYvLYvA1uG1u	290
Db	823	CCAGAAATTCATTCAGACGAAGAAGTACTTGAAGAAATTTATGMAAGGACACAGAA	882
QY	291	AspG1yA1aPhe1leG1yLYvAaEni1eAaBpAaBpA1aThr1a1eA1uA1eA1u	310
Db	883	GAAAGAACTCTTGTGGAAACCTTTACCAAGCAATCTTCTGCTTGGTGGAATTTTG	942
QY	311	1leG1naBpSerSer1eu--LYvLYvAaBpAaBpLYvG1uLYvCyvG1uA1a1eA1u	329
Db	943	ATTCAAGTCATCTCTTACACTCACTCAAGAACTTAACAAGAAATGTAATAAAATCTT	1002
QY	330	G1nLYvSerCyvLYvAaBpProH1eG1uH1eG1uA1eA1uBp1eA1uCyvLYvAaBp	349
Db	1003	GAAAGATAAGTAAAAATTTAAAGAACATGATTTTAAAGGTCTTATCGAGAGATTAT	1062
QY	350	G1LYvSerAaBpAaBpG1yThr1LYvLYvCyvG1uG1uLeuG1naBpRl1eAaBpLYThx	369
Db	1063	AATGCAATATAAGTAAAGAACAAATAATGGAGAACTTAGTACAGATATTGAAGAAACA	1122
QY	370	CyA1yA1lePheThrSerLYvA1aThAaBpAaBpG1eAaBpProThrLYvG1yAaBp	389
Db	1123	TGTAAATTTTTCATTTCAATAAAACCCCTTAGATTCATTTTGGCGATGGAATAAAAT	1182
QY	390	AsnG1u1leA1eA1yTFRG1uG1yLeuPProThrPheLeuSerAaBpAaBpCYvA1aLYs	409
Db	1183	GATGAAATTTATTAATGGGGAAATTTTCAACGTTTCTTAAGCAATAAGATTGACAAA	1242
QY	410	LeuG1uSerTYTCyPheTYrPheG1uLYvA1uCyvAaBpAaBpG1yAaBpA1aCYvLYs	429
Db	1243	TTAGAAATCGATGTCTTTATTTTGAATAAAAGCTGT--AGAAGCGAACTGCATCGAAG	1299
QY	430	Asn1leAaBpA1aThrCYvTYrLYvAaBpG1yLeuAaBpA1aAaBpLYvA1eA1uG1n	449
Db	1300	AATATCGAAGACATCTCAACAGAGAACTTGACACATTAGCAAAATGAAAGTATTCAA	1359
QY	450	G1uAaBpAaBpG1yMeA1eA1eA1yG1SerAaBpAaBpSerTYrPheG1uLYvPheG1n	469
Db	1360	AAAGAAATTCGAGAGAAATGCTGCATGTTCAAAATTAACATGGCTTACTAGTTCCAAAA	1419
QY	470	G1uLeuVAlLYvA1CYvG1uLYvA1uLYvG1uAaBpAaBpG1ySerPheSerAaBp	489
Db	1420	AAACTCTATGAAGGTGCAAAAAAGTAAAAAGAAATAAAGAGATTTTCCAGATAT	1479
QY	490	G1uLeuPhe1leA1uCYvA1eA1nProA1eA1yA1a1aAaBpAaBpThrH1aAaBpA1eA1u	509
Db	1480	GAATATTTTGTCTTATGTGTHCAACATCAAAAGACACTGATTGCTTTGCGATGATCTT	1539
QY	510	ArgMeTYrTh1lePheLeuAaBpG1nLeuAaBpG1nLYvAaBpAaBpPheProThrAaBp	529
Db	1540	CGGATGAAGAACTATCTTTTGGCAAGATTTTGAACGAAGACGAGATTTTCCAGTAAA	1599
QY	530	LYvAaBpCYvLYvG1uLeuG1yAaBpCYvG1naBpAaBpG1yG1naBpSerLYvG1u1le	549
Db	1600	GAAAGCTCGAAGAAATTTAAAGAAATGTGAAGGCTTTAAGAAAGGATTTCTAAAAAAATT	1659
QY	550	ThrTYrProCYvH1aThr1eA1uG1nG1CYvAaBpAaBpG1yThrThrG1u1leA1eA1u	569
Db	1660	GAATGGCCATGTACTATTAAGCAAAATTTGTGATCAATTGGAAGAAACGCAAAAGATTTG	1719
QY	570	LYvG1nVAl1eA1uAaBpG1uH1eLYvAaBpThr1eA1yAaBpG1nG1uSerCYvA1aLYs	589
Db	1720	AAAGAACTTTATCTTAATTAAGAACTAATTAATTTGAATAATCAAGAAATTTGGCAATC	1779

QY	590	YITREILYSGIILYSGYSAENLYETPTSERTRGRNGIYASBPAPRPHSESRPhVal	609
Db	1780	TATTTGAAGGAAATGCATGATGATGGCTTGAAAGAGAAATTAAGACGTTCTCTTTTA	1839
QY	610	CYValPheGlnAenAlaethrCYeGluueneValIySaBPValIySaBPArgCYeGlu	629
Db	1840	TGTGCTTTCGAAATATAGACCTTGCAGATATATGGTAAAGATGTGAAAATCAATGCAA	1899
QY	630	ValPheIyIySaenIIeIySaIaserTrIleIegIuPheIuGluSaenThrAsn	649
Db	1900	ATATTTGAAAAAACATTAATAAAATATCCAGAGTTTGAT-----AGTAAACT	1947
QY	650	IyIeIleThrThrIleGluIuArgAsnCYerProSerTrpHsthTrYrCYaSaenArgPheSer	669
Db	1948	AAATATAGAAAGATTAGGACATATTGCTCTTTTGGCCCACTGCATCGCATAGATTGGA	2007
QY	670	ProAsnCYerProGluIleuThrIyGluIuSaenSerCYethrIyIleIyIyGHIArgGlu	689
Db	2008	CCCAATTGCCCGGACTTTGABAAA---ATTAATGTGAAGCTTGAAAATATATTCGAA	2064
QY	690	ProBheTrYrIyArgIySaIeIeGluIuSaPAlaIeulYsaIeIuIeGluGluIyIyIeu	709
Db	2065	CCTTATATATAGAAAGACCTTGAAAATGCACTTATATTGATTAGATTAGAGACATCTT	2124
QY	710	ThrAspIySerIyCYeGluProAlaIeulYsaIyArgTrYrCYethrValAlaGlyAsnAl	729
Db	2125	GATTAAGAAAAAACTGCACAAATCTTTATATAGTACTGTACACTTGGGATCAACA	2184
QY	730	AsnAsnAlaserIleSerGluCYeIySaIaenThrIySaBPAsnSerGluYlaser	749
Db	2185	GGAAATTAACACTTAAGGTTTTTGTACAGTCTTACTGTATACAT-----	2232
QY	750	AspGluSaPAlaArgIyGluIeCYeGluIySaIeulYyGluValGluGluIuCYs	769
Db	2233	---GAAACATTTAGAGATAAACTTTGGCAAAAACAGTTCAAGCGGTGAAGAAAATGC	2289
QY	770	IySaIaIeulProThrGluIeGluGluProAlaIaIaBPleulYySaBPTrYrIyThr	789
Db	2290	CAAGGATTTATCAAAAGAACTTGAACAAAAATGATTTTGAAGAAAACATTAAGAT	2349
QY	790	TyrGluGluIeulYyIySaArgAlaGluGluIaIaBPleulYySerSerIeuValIeuser	809
Db	2350	TATGAAAAGTAAGAAAGATACAAAAAATGCATGAAAGAAACAAATCTCGTTTTC	2409
QY	810	IeuIleIyIySaenGluSerAsnValSerIySerAsenSerIySaenIySaBPAsn	829
Db	2410	ACAACATAATCAACGATAATTAACAAAAAG-----GTCAAGCTT	2454
QY	830	AlaValSerAsnGluIeGluAspThrTrIyGHIeIyIleIuArgArgIyAl	849
Db	2455	AGTAGCGCTAGTGTCAAGATATTGTA---CATTTTAAACTGTAAAAAGAAATGA	2511
QY	850	IySaBPValSerValThrGluIeGluIaIySaIaPheAspIeuAlaGluValPhe	869
Db	2512	AAA---GTTCAAGTGACAGAAAAGAAACCAAAAGCGTTTGATTGTAGCACTAGCATTC	2568
QY	870	GluArgTrYrValAspIeulYyGluIuArgCYaenIyIeGluIuSerAspCYeArgIleIy	889
Db	2569	AGCTTATATGTAGACTTAAGAAACGTCACCACTTAAGAGATATGGCAATTGGA	2628
QY	890	GluAspCYeIySaBPleuGluGluValCYeIyIyIyIleAsnIySaIeCYaArgAsnIeu	909
Db	2629	AAAGAAATTA---TGTAAGACACAGTCAAGACATAGAAAATATGTTTAAAAATA	2685
QY	910	IySaPheIeGluValIyProGHIeGluThrValThrGluSerThrThrThrThrThr	929
Db	2686	GAACCACTGAAAGTAAGCCACATTAATAAAACATAGACGAACCAACATTAACA	2745
QY	930	ThrThrThrThrValAlaAspProIySaIaThr-----GluCYeIyIySer	944
Db	2746	GTCAACGAAACAGTCAAGAGACGAAAAACATAGAGAGACGAGAAATATGCAAATCT	2809
QY	945	IeuGlnThrThrAspThrTrIyValThrGlnThrSerThrHsthSerThrSerThrIle	964

D	b		CTACGACAAACAGCAGTGGGTCTCAACAAGACCTCAACCATTACGACACTCCAGACT	2865
Oy		965	ThrsrThrIlethrSerlyellethrLeuthrThsrTharArgCySlyspProthLys	984
D	b	2866	ACGCCACAGTTAAGTCATGAATAAACCTACCTCGACGAGCGGTCTTAAGCCTCACAG	2925
Oy		985	CyethrThrGIyglugluAAPapAlagLYAspVallyPProserGIygluAargMet	1004
D	b	2926	TGTACGACAGAGAGGAATGAAAGCAGAGAGGTGAAGCCGAGTGAGGAGCTGAGGATG	2985
Oy		1005	SerGIYrrpamVAlmetArgGLValilleValAlAmetValilleserPmetille	1023
D	b	2986	AGTGGGTGAGTGTGATGACAGGGGTGTATTATGCATCATGATTTCATTGATTT	3042
 RESULT 13 AF038556 12792 bp DNA linear PLN 10-SEP-1998				
LOCUS AF038556				
DEFINITION Pneumocystis carinii f. sp. hominis variant regions of major surface glycoproteins (msg1, msg3, msg4) genes, partial cds.				
ACCESSION AF038556				
VERSION AF038556.1 GI:3560524				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL MEDLINE	PUBMED	98380374	Immun. 66 (9), 4268-4273 (1998)	
79712777				
2 (bases 1 to 12792)				
Wei,Q., Turner,R., Sorial,V., Klvington,D., Angus,C.W. and Kovacs,J.A.				
Direct Submission				
Submitted (12-DEC-1997) CCMD, NIH, Building 10, Room 7D43, MSC 1662, Bethesda, MD 20892-1662, USA				
Location/Qualifiers				
1..12792				
/organism="Pneumocystis jirovecii"				
/mol_type="genomic DNA"				
/specific_host="Homo sapiens"				
/sub_species="hominis"				
/db_xref="taxon:42068"				
<668..3712				
/gene="msg1"				
/note="multicopy gene family is expressed by gene rearrangement"				
<668..3712				
/gene="msg1"				
/note="glycoprotein A, expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"				
/codon_start=1				
/product="variant region of major surface glycoprotein"				
/protein_id="AACJ4980.1"				
/xref="GI:3560525"				
/translation="VARAVAKROYTAGSVDEEVRLLALILKEDSKDDKCEKLKHCKEISLANLTPEQVHEKLDKFDSKKDKCKELKNVEKCGDFTELELVKEKASNLKNDKCTKNEOGLFLBEASPDITKGNDRNKCYONKRDVKAETLAILIGKIKFKDKNSCKKLEIVYCOELSSOWSDMLMKICFDOKVCNDVVKETOOKCSFRKLKTEIKITI				
KEDQLKCPRLYTEBCTFYDESQUNDSLKSELBKCKQKNTITYLSYGFPDIIEEITTLAEVDLGEIYRKAABEGVLVGKPLPAQATLVAVALIDPSLTTRTKNEKKKTLIEDKCNLKEDIILIGLCEDYNANKDKDKCELESTDIETCKFFISKTLMIHFQGDENKDNCIIKGWLNISTLSNKDCCTKLSEYCLVFEKESCRSETACKNIRAAICYRGGLTLANEVLOEKMGMLHGSHNKTLSGFOKKLIIVCKRVKENGVPPSNELFVLCAQPSKAARLVLSHDLPKKTI FLDDILNRKRDPFKRECEBLKCALRKDSKIEMPHCTLSQNCODLRNAKELKELLINENKDIILKQENCGMYLKEKCNERRRNERNRPSLICAIQNTRCRM				





QY	330	GINLYSE	CYVLYA	ASNP	PROHI	EGU	NI	EGU	LA	LEU	GI	USE	LEU	CYVLYS	LYS	ASN	349
		:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	
DB	1670	GAA	GAT	TAA	TAT	GT	AAAA	TTTT	AAAA	AGAA	C	TG	TAT	TAT	AAAA	GGCT	ATG
QY	350	GI	LY	SE	AS	NP	AS	PG	CL	Y	TH	IR	LY	SV	CS	GI	UG
		:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	
DB	1730	AAT	G	C	A	A	A	A	A	A	T	A	G	A	C	T	T
QY	370	CY	SV	IL	EP	HE	TH	SER	LY	SV	AL	TH	R	AS	NA	ARG	LEU
DB	1790	TGT	A	A	A	T	T	T	T	T	T	T	T	T	T	T	T
QY	390	AS	NG	UL	LE	VAL	G	LY	TP	GI	LY	EU	P	OTH	R	PH	LEU
		:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	:::	
DB	1850	GAT	G	A	A	T	T	A	A	T	T	A	A	T	T	A	A
QY	410	LEU	GI	US	ER	TY	CY	AP	HE	TY	SV	LY	SV	CY	P	RO	AS
DB	1910	TTG	A	A	T	T	G	A	T	T	G	C	T	T	T	A	T
QY	430	AS	NI	LE	ARG	AL	ATH	CY	STY	LY	SV	ARG	GI	LY	EU	AS	PA
DB	1967	AAT	A	T	C	A	G	A	G	C	A	T	G	C	T	A	C
QY	450	GI	US	EN	ET	ARG	LY	ME	LEU	NI	EG	LY	SER	AN	LY	SER	TP
DB	2027	AAA	G	A	A	A	T	G	C	G	A	A	T	G	C	A	T
QY	470	GI	US	EU	VAL	LY	SV	AL	CY	SV	LY	EU	LY	SV	GI	US	EU
DB	2087	AAA	C	T	A	T	A	G	A	A	T	G	C	T	A	A	A
QY	490	GI	US	EU	PH	EL	LE	US	CY	VAL	GI	NP	RO	AL	LY	SV	AL
DB	2147	GAA	T	T	A	T	T	G	C	T	T	A	T	G	T	A	T
QY	510	ARG	ME	LY	ETH	IL	EP	HE	LEU	ARG	LI	NG	LE	US	AS	PG	LI
DB	2207	CGG	A	T	G	A	A	A	A	A	A	A	A	A	A	A	A
QY	530	LY	SA	EN	CY	SV	GI	US	EU	LY	ARG	LY	SV	GI	US	EU	LY
DB	2267	GAA	G	C	T	G	C	G	A	A	T	T	A	A	A	A	A
QY	550	TH	TR	P	RO	CY	SE	NI	ETH	IR	LY	UG	LI	NG	LE	US	AS
DB	2327	GAA	T	G	C	C	A	T	G	C	A	T	T	A	A	A	A
QY	570	LY	SG	IN	VAL	LEU	LEU	AS	NP	LI	NI	LY	SV	AP	TH	IR	LY
DB	2387	AAA	G	A	A	C	T	T	A	T	A	T	G	A	A	A	A
QY	590	TY	IR	LEU	LY	RG	LI	US	CY	AS	EN	LY	ET	R	P	HE	SE
DB	2447	TAT	T	G	A	A	G	A	A	A	A	T	G	C	A	T	A
QY	610	CY	SV	AL	PH	EG	IN	AS	PA	AL	ATH	CY	SV	GI	US	EU	LY
DB	2507	TGT	G	C	T	T	G	C	A	A	A	T	A	G	C	T	A
QY	630	VA	LP	HE	LY	SV	AS	NI	LE	LY	SV	AL	SE	TY	IR	LE	GI
DB	2567	AT	T	T	T	G	A	A	A	A	A	A	A	A	A	A	A
QY	650	LY	SV	IL	ETH	TR	LEU	GI	US	EU	LY	ARG	LY	SV	GI	US	EU
DB	2615	AAA	A	T	A	T	G	A	A	T	T	G	C	A	T	T	G
QY	670	P	RO	AS	N	CY	P	RO	GI	LY	EU	LY	SV	GI	US	EU	LY
DB	2675	CCC	A	A	A	T	T	G	C	C	G	A	T	T	T	G	A
QY	690	P	RO	PH	E	TY	IR	LY	SV	AL	LEU	GI	US	EU	LY	SV	AL

LOCUS	DEFINITION	SEQUENCE 3 FROM PATENT US 6664053.	DNA	LINEAR	PAT 20-FEB-2004
Db	2732	CTTATATTAAGCAAGAGACCTTGAATAATCAGCTTATATTGCTTAAAGGACATCTT	2731		
Qy	710	ThraBpYserSerybCYeGluPProAlaLeuBysArgTTCyPthrValAlaGlyAsnVal	729		
Db	2792	GATAAGAAAAAACTCGCAAAACAATCTTGATAGTACTGTACACTGTGGATCAACA	2851		
Qy	730	AsnBsnAlaSerLLeSerGlyLeuCYeLysValaBsnThrLysBspAsnSerGlyLysSer	749		
Db	2852	GGAATATATAACCTTAAAGGTTTTTGTTGTTTAAAGCTTCTACTGATTAACAAT	2899		
Qy	750	AspGluBsnAlaArgLysGluLeuCYeGluLysLeuValLysGluValGluGluGlyCys	769		
Db	2900	--GAAACATTTAAGATATAAAGCTTTGGCAAAACCTAGTTCAGCCTGTGAAGAAAAATGC	2956		
Qy	770	LysAlaLeuProThrGluLeuGlyGlnProAlaAlaBspLeuLysBspTyrLysThr	789		
Db	2957	CAAGATTTATCAAAAGAACTTGAAAAACCAAAAATGATTTTGAAGAAAAACATTAAGAT	3016		
Qy	790	TyrGluGluLeuLysLysValArgAlaGluGluValaMetAsnLysSerSerLeuValLeuSer	809		
Db	3017	TATGAAAAAGTAAAAAGGATATACAAAATATGCAATGGAAGAAACAAATCTGTTTTTCA	3076		
Qy	810	LeuLysLysLysBspGluSerAsnValSerLysSerAsnSerLysBspLysAsn	829		
Db	3077	ACAACTAAATCAACGATTAATTAATAACGAAAAAGCA-----GTCAAGCTT	3121		
Qy	830	AlaValSerAsnGlyLeuGluAspThrThrLysHisValLysLLeuLysArgGlyVal	849		
Db	3122	AGTACGCTACTGATGATTCAGATATTTGTA--CATTTTAACTTGTTAAAGAAATCA	3178		
Qy	850	LysBspValSerValThrGluLeuGluValaLysAlaBspAspLeuValaGluValaPhe	869		
Db	3179	AAA--GTTCAAGTGACAGAAAAAGAACAAACGTTTGATTTGGTACCTAGCATTC	3235		
Qy	870	GlyArgTyrValaBspLeuLysGluArgCysAsnLysLeuGluSerAspCysArgLys	889		
Db	3236	AGCTCTTATGTAGAGTTAAAGAAACGCTGCACATCTAAAGATGATTCGAAATTTAGA	3295		
Qy	890	GluAspCysLysBspLeuGluGluValCysLysLysLysLeuValaCysArgAsnLeu	909		
Db	3296	AAAGAAATTA--TGTAAAGCACAGTGCACAAAGATAGAAAATATGTTTAAAAATA	3352		
Qy	910	LysPProLeuGluValLysProHisGluThrValThrGluSerThrThrThrThrThr	929		
Db	3353	GAAACACAGAAAGTTAAAGCCACATGAATATAAAAAAGTAACGAAACCAATTAACAACA	3412		
Qy	930	ThrThrThrValAlaBspProLysAlaThr-----GluCysLysSer	944		
Db	3413	GTCAACAGAAACAGTCMAAGAAAGCAAAAAACAATAGAGACGAGAGAAATGCAAACT	3472		
Qy	945	LeuGlnThrThrAspThrThrThrValThrGlnThrSerThrHisThrThrThrThrThr	964		
Db	3473	CTCAGCAACAAGACAGCGGGTCAACAAGACGTCAACCCCTTACACAGACTCCACACT	3532		
Qy	965	ThrSerThrLysThrSerLysAlaThrLeuThrSerThrArgArgCysLysProThrLys	984		
Db	3533	ACGTCCACAGTTACGTCAAGATTAACACTGACCTCGACGAGCGGTGTAAACCTACGAG	3592		
Qy	985	CysThrThrGlyGluGluBspAspAlaGlyAspValLysProSerGluGlyLeuArgMet	1004		
Db	3593	TGTACACAGAGAGAGAAATGATGAAGACGAGAGAGTGAAGCCGAGTGAAGGCGCTAGAGATG	3652		
Qy	1005	SerGlyThrPanValMetArgGlyValLLeValaLMetValLLeSerPheMetLle	1023		
Db	3653	AGTGGGTGAGATGTGATGAGAGGGGTGTTATTACCAATGATGATTCATTCATGATTT	3709		

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3006)  
AUTHORS Kovacs, J. A., Huang, S., Masur, H., Fischer, S. H., Gill, V. J. and Mei, Q.  
TITLE Identification of a region of the major surface glycoprotein (MSG)  
JOURNAL Patent: US 6664053-A 3 16-DEC-2003;  
FEATURES location/Qualifiers  
source 1. 3006  
/organism="unknown"  
/mol\_type="genomic DNA"

## ORIGIN

## Alignment Scores:

Pred. No.:	9.1e-146	Length:	3006
Score:	2404.50	Matches:	479
Percent Similarity:	65.37%	Conservative:	195
Best Local Similarity:	46.46%	Mismatches:	320
Query Match:	44.44%	Indels:	37
DB:	6	Gaps:	19

US-10-654-416-14 (1-1023) x AR438435 (1-3006)

QY 1 MetAlaATgAlaValLyAspGlnAlaAagLyThrGlnAsnSerTLeAspGluGluHis 20  
DB 1 GTGGCGCGGCGGTCAACGGCGGCTGCA--GCACAGAAATGAGTTGAAGAAATAT 57  
QY 21 ValLeuAlaLeuLleLeuLysGluAspGlyLeuSerGluGlnGlyCysValLeuLeu 40  
DB 58 CTTTGGCTTGATTTGATTTGAAATGAGTATGAAATTAATGATTAATGTAAGTAAAGTTG 117  
QY 41 LysLysTyGCGInGluLeuThrGlnAlaLysLeuAsnLleGluGlnValHisArgLys 60  
DB 118 AAAGGATTTGTGGAAGTTTAAATAAATGTAACAAAGAAACCAAAAGAACTGAAGAAAG 177  
QY 61 LeuLysGlyPheCysGluAspGlyLysAla--AspThrLysCysLysGluLeuLysAla 79  
DB 178 TTAGACGGAATCTGCAAGATGATTAACAATAGAACAAATGCAAGAAATGCAAGAAACA 237  
QY 80 AsnLleGluLysCysThrThrLysGlyLysLeuLysGluValLleLysLysLys 99  
DB 238 AAGGTTAAAGCAAGATGTAAGTTTTCACACAGAACTGATGAAGCTCAAAAGGGA 297  
QY 100 IlegInLleLleThrAspLysAspCysLysGluAsnGluGlnGlyCysLeuPheLeuGlu 119  
DB 298 GCTTCAACATTGAAGATPATGATTTGTAAGAAATGAAACGACATGCTGTTTGGAG 357  
QY 120 GlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuAspAsnLysCysTyGln 139  
DB 358 GGAGCATGTCCAAACGAACCTTAAGATTAATGAACTGACGAAATTAATGTTATCAA 417  
QY 140 LysLysAspAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsn 159  
DB 418 AAAAACAAGACGACGTAGCAAGAAAGCTCTTTTAAGAGTACTTAAGAGGGAACCTTAAG 477  
QY 160 GlySerValLleCysGluLysLysLeuLysGluLleCysProValMetGlyArgGluSer 179  
DB 478 GATAAAACACATGCAAAATAAGTTAAAGGGGGTTTGTCAAGATTCACAAAGAAAGT 537  
QY 180 AspGluLeuThrAsnLysCysLeuAsnGlnLysGluThrCysLysAsnLleLeuLleGlu 199  
DB 538 GATGAGCTAAATTAATTTATGCTTGAAGAAAGAAAGCTGTGAGATCTGTATCTAAG 597  
QY 200 LysAspLysCysGlyThrLeuLysThrAspValSerAlaLleGluLys-----Ser 217  
DB 598 AAAAGATTCAAATGCAAACTCTCAAGAGAAAGAAATGATCTAGCTTGAAGAAAGAGAT 657  
QY 218 PheLysLysGluThrCysLeuGlnLeuLeuGlnGluGlnCysTyPheTyLleGlyAsnCys 237  
DB 658 TTATTTAAAGAAATGTTTATTTATTTCTTGAAGATGTTACTTTTATGGGTCAAAGCTGT 717

QY 238 GlyAspAspLleLleLysCysLleGluLeuGlyLysCysGlnGluGlnAsnLle 257  
DB 718 ---GAACAGATCAGCCAAAGTGTAAAGATTTGCCAAGCAAAATGTCAAAAGAAATCTC 774  
QY 258 AlaTyMetCProPogLyProAspPheAspProThrArgProGluAlaThrLleAlaGlu 277  
DB 775 GTTATTCAGACACAGGTTCACTTTGATCTTACGAAATTAAGATTAAGATTTAGCTTACAGAA 834  
QY 278 AspLleGlyLeuGlnGluPheTyLysLysValGluGluAspGlyValPheLleGlyLys 297  
DB 835 GAAATAGACCTAGAAATAATGTATGATGAGCAGTAGAAAGAAAGAAATTCATATTGGAAG 894  
QY 298 AsnHisLeuArgAspAlaThrAlaLeuLeuAlaLeuLleGlnAspSerLeuLys 317  
DB 895 CCATCAATTAAGAGTGAAGCTGCTTATTTGCGATTTATTAAGCAAGAGTATGCT----- 948  
QY 318 LysLysAspAspLysGluLysCysGluGluLalaLeuGlnLysSerCysLysAspProHis 337  
DB 949 CAAATATCTTTAAAGATCAATGTGAAGATCTTTTAAABAAAATGTGAAACCTTTAA 1008  
QY 338 GluHisGluAlaLeuGlnLysSerLeuCysLysAsnGlyLeuSerAsnAspGlyThrLys 357  
DB 1009 GAGCATATTATTATAAAGATTTATGATGATTAAGACTATCAGCATTAATCCAAAGAA 1068  
QY 358 LysCysGluGluLeuGlnAsnAspLleAsnLysThrCysLysLlePheThrSerLysVal 377  
DB 1069 AAATGCCAAGAACTTAATTAAGAGTTA--ACAACCTATTTTAACCTTTCTTCAA 1122  
QY 378 ThrAsnAsnArgLeuPheAspProThrLysGlyAsnAsnGlnLleValGlyTrpGluGly 397  
DB 1123 AGGATTGAGAAATTTTCGCTCCAGCT--AATGTAAGGAAATTAATGCTTGGCATATG 1179  
QY 398 LeuProThrPheLeuSerAsnGluAspCysAlaLysLeuGlnLysSerTyCysPheTyPhe 417  
DB 1180 TTGCATACATTTCTTGATGAAGAGAGTGTACGAAAGCTGTGCGATGTTTATATTG 1239  
QY 418 GluLysLysCysProAspGlyGluAsnAlaCysLysAsnLleArgAlaThrCysTyLys 437  
DB 1240 AAAAGCCAGCTCA--CTTGAAGACCCCTGCAATTAATTAAGCAGCATGTTATATA 1296  
QY 438 ArgLysLeuAspAlaArgAlaAsnLysValLeuGlnGluAsnMetArgGlyMetLeuHis 457  
DB 1297 AAAGGCTTGAAGCAGTAGCAAAATGAACATTAACAAGTAAAGTGAAGGAAATTCGA 1356  
QY 458 GlySerAsnLysSerTrpLeuGlnLysPheGlnGlnGluValLysValCysGluLys 477  
DB 1357 GGTCAATTAAGACATGCTTGAACCTTCAAAAACCTGTGTAAGTGTGCAAAAG 1416  
QY 478 LeuLysLysGluAsnLysGlySerPheSerAsnAspGluLeuPheLleLeuCysValGln 497  
DB 1417 ACGAAAGGAGAA-----AGTATGAATTAATTTGATCTATGATGAAC 1458  
QY 498 ProAlaLysAlaAlaArgLeuLeuThrHisAspLeuArgMetLysThrLlePheLeuArg 517  
DB 1459 CCAATTAAGAAAGGCTCTTACAGTGTCAACAGATTTGCGAATGAGGCGACTGTTGCA 1518  
QY 518 GlnGlnLeuAspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArg 537  
DB 1519 GAGCATTTGAAGCAAAAGCAGATTTTCCACAGAAAGATGTAAGATTAAGATTAAGAA 1578  
QY 538 LysCysGlnAspLeuGlnLysAspSerLysLleThrTrpProCysHisThrLeuGlu 557  
DB 1579 AAATGTAGGCTTAAAGAAAGATTCAGAGAAATTAATGATGCTATTAACGTTAA 1638  
QY 558 GlnGlnCysAsnArgLeuGlyThrThrGluLleLeuLysGlnValLleLeuAspGluHis 577  
DB 1639 CAGCATTCATTCGGCTGAAGACATAGACATTAAGAGAGACTTCTTAAGAAAT 1698  
QY 578 LysAspThrLeuLysAspGlnLysSerCysValLysTyLysGluLysCysAsnLys 597  
DB 1699 AAAGGATTTTAAAGATGAATAAGTGTCAAGAAAGAAAGTGAAGAAAGATGTGAAAA 1758  
QY 598 TrpSerArgArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCys 617

```

Db      1759 TGGTTTGAAGAGAAAATAATAATTTTTCGGCTTGTCTGACCTTGGAACCTGTTTGC 1818
Qy      618 GlnuemetValIySaapValIySaapArgCysgluValIyPheIySylSaenIIySaIa 637
Db      1819 AAAAAAGATCACTAATAAGTTGAATCTTAATGTAATATTGAAGAACAATGGAACCT 1878
Qy      638 SerTyrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 657
Db      1879 ATGAACCTTATTAAGTGAATAGCTAAAAAAGAGAAAAAATTA----- 1920
Qy      658 CysProSerTyrPheIySylSaenIyPheSerProaenCysProIyIleuThr--- 676
Db      1921 TGTGAATTTTGGCTCCATATTGTAAAAAGTACAGCAAAATTTGTAAAAAATCTTAAAAAC 1980
Qy      677 -----LySgluAaenSerCysThrIyIleIleIleIleIleIleIleIleIleIleIle 694
Db      1981 GAGAGAAAAGATGGGCAATGCAAAAACCTCAATAAAAAGTCAAAATCATTCCTTGA AAAA 2040
Qy      695 LySaIaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 714
Db      2041 GAAGCTTATGAATAATAAGTTGTAGAAAGATTGAAGAGTATTAACAACGTAGAGAA 2100
Qy      715 CysgluProaIleuIySaargTyrCysThrValIaIleuIleuIleuIleuIleuIleu 734
Db      2101 TGTAAACAATCACTTAATATATATACGTACCAATGAAAAAGCAGAGAAAT---GGGTTG 2157
Qy      735 SerCysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 754
Db      2158 GAAACTTTGTGCAAA-----AGCAAAAGAAAACAC---AAGAGTGCATTAAAGTTAGA 2208
Qy      755 LySgluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 774
Db      2209 GAAGAACTCTGTGAAAACTTAATAAAAGCTTAATAAAAGAAAATGCTCAAAATTGAAGAC 2268
Qy      775 GluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 794
Db      2269 GAGCTTGAGAAAGTAAAGAGCTCTTGAAGAAAAGAAAGTAAATAAAATAATAA 2328
Qy      795 LySaargIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 814
Db      2329 GAAGAAACAGAAAAAGCCATGGAAGATGCAAACTTATTTATTCAGAGACGAAAGGAACT 2388
Qy      815 GluSerAaenValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 834
Db      2389 GATATATTAATTAATTAAG-----TCAGTAAATTAAGACTCATCTGATACCTTAAGGAA 2442
Qy      835 LeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 854
Db      2443 GGAAGAAAGCACAACA-----GCAATTAACTGTAAAGAAATGCAAAA---GTGCATGTA 2496
Qy      855 ThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 874
Db      2497 ACAGAAAAAGAAATTAAGCAAGCATTTGATTGGTACAGAGAGATTTGATCTCTATCTAGAA 2556
Qy      875 LeuIySgluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 894
Db      2557 TTGAAAGAAATTAATTAATCACTTCACTGAAAGATTTGGTTTCAAAAAGAGAGT---GAC 2613
Qy      895 LeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 914
Db      2614 TGTGAAGATTCATTTAAAAAGATACAGGGAATATGTTCAACATTAGAGCCACTAAAGTG 2673
Qy      915 LySProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 932
Db      2674 AGACACACACGAAATTAAGTAAACATAAACATAACACACACACACACACACACAACT 2733
Qy      933 ThrValIaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 952
Db      2734 ACCTTTAAAGCGCAAGGCAACAGACTGCACTTTTACAGACAAAGATACGAGGCTC 2793
Qy      953 ThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 972

```

```

Db      2794 ACAAAAGCTGAGCCCATCTAGACATCCAAACCAATCTTACAGTCAAGATA 2853
Qy      973 ThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 992
Db      2854 ACCTTGACCTTGCAACACCGGTGTAGCTTACGAAGGTGTACAGACAGAGAGAAAGATGAA 2913
Qy      993 AlaIySaapValIySaapIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1012
Db      2914 GCAAGACACGTGAACCAAGATGAAGGTTGAGATGAGTGAAGTGAAGTGAAGTGAAGGCGG 2973
Qy      1013 ValIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 1033
Db      2974 GTGTTATTAGCAATGACGATTTCAATCATGATTT 3006

RESULT 15
PMCSGINS
LOCUS      3563 bp      mRNA      linear      PLN 30-JUL-1999
DEFINITION Pneumocystis carinii mRNA for major surface glycoprotein, complete
            cds.
ACCESSION D21827
VERSION   D21827.1 GI:425784
KEYWORDS  major surface glycoprotein.
SOURCE    Pneumocystis carinii
           Pneumocystis carinii
           Burkholderia; Fungi; Ascomycota; Pneumocystidomycetes;
           Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 3563)
AUTHORS   Kikada, K., Wada, M. and Nakamura, Y.
TITLE     Multi-gene family of major surface glycoproteins of Pneumocystis
           carinii: full-size cDNA cloning and expression
JOURNAL   DNA Res. 1 (2), 57-66 (1994)
MEDLINE   96051981
PUBMED    7584029
REFERENCE 2 (bases 1 to 3563)
AUTHORS   Kikada, K.
TITLE     Direct Submission
JOURNAL   Submitted (27-OCT-1993) Kazuhiro Kikada, Institute of Medical
           Science University of Tokyo, Department of Tumor Biology; 4-6-1
           Shirokanedai, Minato-ku, Tokyo 108, Japan
           (E-mail:kikada@hgc.ims.u-tokyo.ac.jp, Tel:03-3443-8111(ex.308),
           Fax:03-3443-6319)
           Submitted (27-Oct-1993) to DDBJ by:
           Kazuhiro Kikada
           Department of Tumor Biology
           Institute of Medical Science University of Tokyo
           4-6-1 Shirokanedai, Minato-ku
           Tokyo 108
           Japan
           phone: 03-3443-8111 x308
           Email: kikada@hgc.ims.u-tokyo.ac.jp
           Fax: 03-3443-6319.

FEATURES
            source
            1..3563
               Location/Qualifiers
               1..3563
                  /organism="Pneumocystis carinii"
                  /mol_type="mRNA"
                  /strain="IMS-PC84"
                  /db_xref="taxon:4754"
                  /dev_stage="trophozoite"
                  333..3563
                     /codon_start=1
                     /product="major surface glycoprotein"
                     /protein_id="BAA04851.1"
                     /db_xref="GI:425785"
                     /translation="MAAPVKGAVQGAADIDIEKHLLAFKAEKYKBNCKOBLKY
                     CBEIKKIDGSDVNVKNGICDEGKQDKLGEVAKVLAKEGELQELAKDKIDN
                     CEEYKEKICILIEDHGAIKENCVNLKEKYKLRKREKVAELLFRLAGSAKDQSK
                     EKIEKVCPLRSSEIEIMSFCLNPSGTGCEGLKGLGVQPLKEDLDGELAEKCHER
                     LEKHFPYGAENAKKCDPEKCKKNNITTKAPESDSPKPSALRLRIGPDDVYK
                     AKRGHILIGTGVLDLPKSGTNFLDILLVLSRDEKSTDDKCTALAEKCDASKYLQ
                     HNLKELCDKDKKCKELLDVNVKESCTNLKLYIKDLSRFBKAEKSDLSKQGL
                     PLTFYGECAELESECFYLENACKDNEIGACONLRACYKKQDRLNLFQLEELG
                     NLGIYRSDPGCKKXVYVNCYKLDKCYLPRLCYPELICYGSNDIFLOSKEISALL
                     DDGRDFLEKDCVELKEKCDLSDSLNLNKECITTLKRRCSEYFKVTERPRKVFLEKDD

```

## ORIGIN

DSLMTQENCTVLAHEKCTLYKRRKNSVSICALPEETCSMVVHTSODCKYIKENIK  
 NGEIVGKIEKANGNEALIEELCTTWGRHQJLVENCPEQLKKEGNNNNCALDEKCK  
 ETFOKLEIDELTHLKGSLKDEIECKTKLEKPEBKNEKLAFTLLKCDNDKGNV  
 CTKLVDKVKCKPTLKTDLBEAKKELKTKREKEDALCKAAKSTKEXKLISKSGKXA  
 MPTANGSDSTLVPPPOAPAGSSGSGSPVPPPPGPPGPPGPPGPPGPPGPPGPPG  
 TGOIKNNAKKGLIKRAVVDGVSFAVVAFPAATVIALEMVLEKESGALOLDGPFRE  
 ECGSKQACEKIDELCGSIKPELIKPHHTFOKEISTTTTTTTTTTTTTTTTTTTTT  
 TTTTPGSGKVTIECTMTQTTDWTWTSLSHTSTSTSTSTSTSTSTSTSTSTSTST  
 TTDSNKETDKGEGEBBDVKIPNDKIRVPMKIRIMLGLVIMGM"

## Alignment Scores:

Pred. No.:	2,82e-102	Length:	3563
Score:	1731.50	Matches:	399
Percent Similarity:	53.42%	Conservative:	203
Best Local Similarity:	35.40%	Mismatches:	370
Query Match:	32.00%	Indels:	155
DB:	8	Gaps:	32

US-10-654-416-14 (1-1023) x PMCMISIMS (1-3563)

QY 1 MetAlaArgAlaValLysArgGlnAlaAa--GlyThrGlnAsnSerIleAspGluGlu 19  
 Db 333 ATGGACGCGCGGTTTAAGAGCAACAGTACAGAGGAGGAGGATGATGATGAGAA 392  
 QY 20 HisValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluCysLysLysLys 39  
 Db 393 CACCTTTGGCTTCATTCGAGAGAGAAATATAAGAAAGAAATTAATGCAACAAANA 452  
 QY 40 LeuLysLysThrCysGlnGluLeuThrGlnAlaLysLeuAsnIleGluGlnValHisArg 59  
 Db 453 CTGGGAAATATGTGGAAGATTAAAGAAATATAGT---GGTGGTCCGATGTGATATAA 509  
 QY 60 LysLeuLysGlyPheCysGlyLysAspGlyLysAlaAspThrLysCysLysGluLeuLysAla 79  
 Db 510 AATGTTAAAGGACTTGTGAGATGAGAAACAACAGATTAATGCAA---CTGAAGGC 566  
 QY 80 AsnIleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLys 99  
 Db 567 GAAGTTGAAAAAGATTAAGAACTTTTGAAGGGGAACTTCAAGAAAGCATTTGAAAGAC-- 623  
 QY 100 IleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGlu 119  
 Db 624 -----ATTAAGATGAAATGTAAATGTAAATGTAAATGTAAATGTAAATGTAAATGTAA 674  
 QY 120 GlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysEtyGln 139  
 Db 675 GAGACGATCATGCGTATTAAAGAGAACTGTCAACTTGAGGAGAAATGTTACAA 734  
 QY 140 LysLysAspAspLysValAlaGluGluValLeuLeuArgAlaLeuLysSerAspLeuAsn 159  
 Db 735 TTGAAGCCTGAAAGGTGAGAGAGGCTTTCTTTCAGGGCACCTGGAGGGGATGCTAAA 794  
 QY 160 GlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyAspGluSer 179  
 Db 795 GATGATGTTAAATGTAAAGAAAGATAGAAAGGTGGCCACTGTATACCGGAAAGAGC 854  
 QY 180 AspGluLeuThrAsnIleCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGlu 199  
 Db 855 GAAGAAATGATGCTCTTTTGGCTTAATCCAGTGAAGTGTGAGAGACTGAAGAGAAA 914  
 QY 200 LysAspLysLysCysGlyThrLeuLysThrAspValSerAlaLeuGlySerPheLys 219  
 Db 915 TTGGGTACTGTTTGGCAGCTTTAAAGAAAGATCTTAA-----GATGGCGAATTA 965  
 QY 220 LysGluThrCysLeuGluLeuGluGlnCysEtyrPheTyrlleGlyAsnCysGlyLysP 239  
 Db 966 GCGGGAAGTGTGAGAAAGATTTAAGATTTTACCGGAAAGCGGTAAATAT 1025  
 QY 240 AspAspIleIleLysCysIleGluLeuGlyGlyLysCysGlnGluGlnAsnIleAlaTy 259  
 Db 1026 -----GCGAAATGTAAAGACTTTGAGGAGAAATGTAAAGAAATTAATTTACATAT 1076

QY 260 MetProGluGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIle 279  
 Db 1077 AAACACCGGAATCCGATTTTAAAGTCAGTGAACCGAGTGGCTGTTTAAGAAAGATT 1136  
 QY 280 GlyLeuGluGluPheTyrlLysLysValGluLysAspGlyValPheIleGlyLys----- 297  
 Db 1137 GGGTTATATATGCTATTAAGAAAGCTGAAACATAGGATTTATTTGGAAGAAACAGGA 1196  
 QY 298 -----AsnHisLeuArgAspAlaThrAlaLeuLeuAla 308  
 Db 1197 GTGATTTTAAACAGAAAGTCAGTAACAATTTTCTGCAAGAT-----CTCTTGCTA 1247  
 QY 309 LeuLeuIleGlnAspSerSerLeuLysLysAspAspLysGluLysCysGluGluAla 328  
 Db 1248 GTCTTGACAGAGATGAA-----AAGGACGAGATACAGTAAAGATGCACTTAAGCG 1301  
 QY 329 LeuGlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluLysLeuCysLysLys 348  
 Db 1302 TTGAAAGAAATGATGCTCTTCAAGATTTGATGATGATGATGATGATGATGATGAT 1361  
 QY 349 AsnGlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsn--AspIleAsn 367  
 Db 1362 -----AATGATTAACAAATTAATGCAAGAAATTAATGCAAGATGTAATGTAATAA 1409  
 QY 368 LysThrCys-----LysIlePheThrSerLysValThrAsnAsnArgLeu 382  
 Db 1410 GAAGATGTCAAAATCTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1463  
 QY 383 PheAspProThrLysGlyAsnAsnGluIleValGlyTrpGluGlyLeuProThrPheLeu 402  
 Db 1464 TTTGAAAGCTGAA---AATCAAGATCTTTATCGTGGGACAGCTTCCAACTTATTT 1520  
 QY 403 SerAsnGluAspCysValLysLeuGluSerTyrcysPheTyrcysGlyLysLysCysPro 422  
 Db 1521 ACAGAGGAGAGTGGCAGAACTTGAAGTGGTCTTATTAAGAAATGCGCTGTAAA 1580  
 QY 423 AspGlyGlu-----AsnAlaCysLysAsnIleArgAlaThrCysTyrlLysArgGlyLeu 440  
 Db 1581 GATATATGATGTGTGAAGCGTGTCAAAATCTCAAGATCAAGCTGTATTAAGAAAGGACAA 1640  
 QY 441 AspAlaArgAlaAsnLysValLeuGlnLysAsnMetArgGlyMetLeu----- 456  
 Db 1641 GACAGGATGTTGAATTAAGTCTTTCAAAGAAATTAAGGCGAAATCTTGCTGTAA 1700  
 QY 457 HisGlySerAsnLysSerTrpLeuGluLysPheGlnGlnGluLeuValLysValCysGlu 476  
 Db 1701 TATGTAAGCATCTCGTGTATTTAAATAATAT-----GTAAGTGGGAAAGCTGTACA 1751  
 QY 477 LysLeuLysLysGluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysVal 496  
 Db 1752 AAACCTTGATTA-----AAATATCTTCCAGAGATGCTT 1784  
 QY 497 GlnProAlaLysValAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeu 516  
 Db 1785 TATCTTAAGAACATATGTATAGCGCTTTCAAATGATATTTTCTTCAATCCAAAGAGTTA 1844  
 QY 517 ArgGlnGlnLeuAspGlnLysAspAspPheProThrAspLysAsnCysLysGluLeuGly 536  
 Db 1845 AGTCGCTTTGATGATCAAAAGGAGATTTTCAATTAAGAAAGATGTTGTTGAGTTGAAG 1904  
 QY 537 ArgLysCysGlnAspLeuGlyLysAspSerLysGluIleThrTrpProCysHisThrLeu 556  
 Db 1905 GAGAGGTGTGAACCTTAAGTATGATTAATTAAGAAAGATGATTAACATTTG 1964  
 QY 557 GluGlnGlnCysAsnArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGlu 576  
 Db 1965 AAAAGACCTGTGAATACTTTAAGTTTACAGAAAGATTTAAGAAAGTATTTTGAAGAAA 2024  
 QY 577 HisLysAspThrLeuLysAspGlnLysCysValLysTyrlLysLysGluLysCysAsn 596  
 Db 2025 AAGATGATGCTTAATGATCAAGAAAGATGTCAAAAGATTTGATGAGAAAGATTAAT 2084  
 QY 597 LysTrpSerArgArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThr 616



*This Page Blank (uspto)*







[illegible]

QY	881	lysleuengluserarpCyysarg11lelygluabspCyelsysarpleuugluValCylslyg	900
Db	2641	AAATTTGAATCAGATTGACGAAATTAAAGAGGATTGCAAGACTTAGAAGAGTATTCAAA	27000
QY	901	LyslleasnylsalCyBargAsnleuLyBProleuGluValLyBProHISgluThrVal	920
Db	2701	AAGATTATTAAGGTTGTGTCCGAATCTGAAGCTCTGGAGGTGAAGCCGACGAAACAGTG	27600
QY	921	ThrGluserThrThrThrThrThrThrThrThrThrThrValAlaaspProLyAlaThr	940
Db	2761	ACAGAAAGTACAAAGACAACTACAAACAACAACAAACCCGTCGCCGATCCGAAAGCAACG	28200
QY	941	GluCyelysSerleuGlnThrThrAspThrThrValThrGlnThrSerThrHisThrSer	960
Db	2821	GAATGGCAAAATCTTACACAGACAACACACATGGGTTCACACAGACATGACACACAGC	28800
QY	961	ThiSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgCys	980
Db	2881	ACGCTTACTACATCATCTACCATCATCAAAAATTAACATTGACATCAACAGGCGATGC	29400
QY	981	LysBProThrLyCySerThrThrGlyGluGluAspAlaGlyAspValLyBProSerGlu	10000
Db	2941	AAACCAACAAGTGTACACACAGAGAGAGATGATGCAGAGACGTAACCAAGTATAG	30000
QY	1001	GlyleuArgMetSerGlyTTPAsnValMetArgGlyValIleValAlaMetValIleSer	10200
Db	3001	GGGTGTGAGAGATGGGTGGAAATGTATGATGAGGGGGGTGATGATGACATGTTATTTCC	30600
QY	1021	PheMetIle 1023	
Db	3061	TTTCATGATT 3069	
RESULT 2			
ID	AA294068	standard; DNA; 3054 BP.	
XX	AA294068;		
XX	15-SEP-2003 (revised)		
DT	05-JUN-2000 (first entry)		
XX	Pneumocystis carinii major surface glycoprotein gene HMSG33.		
DE	Major surface glycoprotein; MSG; HMSG33; human; pneumonia; diagnosis; ds.		
KW	Pneumocystis carinii; sp. f. hominis.		
OS			
XX			
FN	Key	Location/Qualifiers	
FT	CDS	1..3054	
FT	FT	/*tag= a	
FT	FT	/product= "HMSG33"	
FT	FT	/note= "a nucleic acid comprising residues 2887-3132 of	
FT	FT	this sequence is specifically claimed in Claim 28"	
XX			
PN	WO200009760-A1.		
PD	24-FEB-2000.		
XX			
PE	17-AUG-1999;	99WO-US018750.	
XX			
PR	17-AUG-1998;	98US-0096805P.	
XX			
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;		
XX			
DR	WPI; 2000-206025/18.		
DR	P-PSDB; AAY79170.		
XX			
PT	Detection of the presence of Pneumocystis carinii in specimens by		
PT	identification of major surface glycoprotein (MSG) gene sequences using		
PT	two or more oligonucleotide primers derived from human P. carinii MSG		
PT	protein encoding sequence.		

XX Claim 27, Page 87-91; 110pp; English.

PS The present sequence is that of the novel *Pneumocystis carinii* sp. f.  
 CC hominid gene, HMG33, which encodes a major surface glycoprotein (MSG,  
 CC see AAY79170). The gene was isolated by PCR amplification of DNA taken  
 CC from an autopsy lung sample of an HIV-infected patient with *P. carinii*  
 CC pneumonia. It is 1 of 7 novel, claimed genes (see AAY79063-69) of the  
 CC invention that encode human-*P. carinii* MSGs (see AAY79165-71). The MSGs  
 CC include a highly conserved C-terminal region of approximately 100 amino  
 CC acids; this region (also claimed) corresponds to residues 2887-3132 of  
 CC the present sequence. Direct detection or amplification of human-*P.*  
 CC *carinii* MSG-encoding genes, especially by PCR using primers directed at  
 CC the conserved region of the genes, provides a sensitive and specific  
 CC technique for the detection of *P. carinii*, and the diagnosis of *P.*  
 CC *carinii* pneumonia, especially in biological specimens (e.g. blood,  
 CC sputum) from immunocompromised patients such as those with HIV infection.  
 CC (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 3054 BP; 1222 A; 449 C; 670 G; 713 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0 Length: 3054  
 Score: 4398.50 Matches: 838  
 Percent Similarity: 88.81% Conservative: 75  
 Best Local Similarity: 81.52% Mismatches: 98  
 Query Match: 81.29% Indels: 17  
 DB: Gaps: 9

US-10-654-416-14 (1-1023) x AA294068 (1-3054)

OY 2 A1AATGAlaValLyArgGlnAlaAlaGlyThrGlnAenSerIleAepGluGlu---His 20  
 Db 1 GCGCGGGCGGCTCAACGCGAGGTAAACAGAGCAGTCAGGCAATATGATGAACTGCAAT 60  
 OY 21 ValLeuAlaLeuIleLeuLyArgGlnAepGlyLeuSerGluGlnGluCysLyArgLeu 40  
 Db 61 ATTTGGCGCTGATCTCAACAGAAAGATGCATAGAAAGATCAAAATGCAAAAAGTTTA 120  
 OY 41 LybLybLybLybGlnGluLeuThrGlnAlaLybLeuAenIleGluGlnValHisArgLyb 60  
 Db 121 GAAAAATCTCGCAAGAGTTGAAAAAGATCATCAGCATGAGAAAGTAAAGTCAATAAATG 180  
 OY 61 LeuLybGlyPheCysGluAepGly-----LybAlaAepThrLybCysLybGluLeu 77  
 Db 181 CTTAAAGATTTCTGTGGAATATGGAAGCAAGTAAAGCAAAATATGCTCAAGTCTTA 240  
 OY 78 LybAlaAenIleGluLybCysThrThrIleLyb---GlyLybLeuLybGluAlaIle 96  
 Db 241 CAAGCCAAAGTTACGAGGAATGTACAAATTTTAAACACAAAGCTAGACCAAGCTTA 300  
 OY 97 LybLybLybIleGlnIleIleThrAepLybAepCysLybGluAenGluGlnGlnCysLeu 116  
 Db 301 ACAATATCCA-----TCAGATGATATTCACAAAGAGAGGAAAGCAATGCTTA 348  
 OY 117 PheLeuGluGlyValCysSerLybGluLeuLybAepAepCysAenThrLeuArgAenLyb 136  
 Db 349 TTTTGGAGGAGCATGC---CATATCTTGTGTAAGATTTGTACAACTAAGGATCTTA 405  
 OY 137 CysTyArgInuLybLybArgAepLybValAlaGluGlnValIleLeuAenArgAlaLeuArgSer 156  
 Db 406 TGTTAACGAAAAACGTACCGAGATAGCAGAAAGATCTTTTAGGGCACTTGTGT 465  
 OY 157 AepLeuAenGlySerValIleCysGluLybLybLeuLybGluIleCysProValMetGly 176  
 Db 466 GATCTCAATAAAAACAGAAACACATGAAAAAAACGAAAGAGATTGGCCAGCTTTGCAG 525  
 OY 177 ArgGluSerAepGluLeuThrAenLeuCysLeuAenGlnLybGluThrCysLybAenIle 196  
 Db 526 AGGGAAGATGATTAACGAGACTTGTGTGAACACGAAAAAAGACGCGCAGATATTT 585  
 OY 197 LeuIleGluLybAepLybLybCysGlyThrLeuLybThrAepValSerAlaIleLeuGly 216  
 Db 197 LeuIleGluLybAepLybLybCysGlyThrLeuLybThrAepValSerAlaIleLeuGly 216

Db 586 ATAAAAAGAAAAATAAAAATGCATCTTTAAACCAATGTTGCACAGCACTTGA 645  
 OY 217 SerPheLybLybGluThrCysLeuGluLeuGluGlnCysTyArgPheTyIleGlyAen 236  
 Db 646 AGTTTAAAAAAGAAATATGCTTGAATTAATCTTGAACATCTCTATTTTACATTGGAAT 705  
 OY 237 CysGluLybAepAepIleIleLybCysIleGluLeuGlyGlyLybCysGlnGluAen 256  
 Db 706 TGCAGACACGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 765  
 OY 257 IleAlaTyMetProProGlyProAepPheAepProThrArgProGluAlaThrIleAla 276  
 Db 766 ATGTTTATATATACACACAGACCCGATTTTGATTCACACTAGACAGAGCTACTAGCA 825  
 OY 277 GluAepIleGlyLeuGluGluPheTyLybLybValGluGluAepGlyValPheIleGly 296  
 Db 826 GAGGACATAGACCTGATGACCTTATTAATAAAGCAGAAAGAGATGGTGTATTATTTGA 885  
 OY 297 LybAenIleLeuAenAepAlaThrAlaLeuLeuAlaLeuLeuIle---GlnAepSerSer 315  
 Db 886 AAACATCATTTAAAGATGCGACAGCTTTATTTGACGTTGTGTTAAGAAAGATGATACA 945  
 OY 316 LeuLybLybLybAepAepLybGluLybCysGluGluGlnAlaLeuGlnLybSerCysLybAen 335  
 Db 946 GGAATAAT 1005  
 OY 336 ProHisGluHisGluAlaLeuGluSerLeuCysLybLybAenGlyLeuSerAenAepGly 355  
 Db 1006 TCTCACAGCATGAAAGCTCTTAAAAATTTATGTAATATATATATATATATATATATAT 1065  
 OY 356 ThrLybLybCysGluGluGluLeuGlnAenAepIleAenLybThrCysLybIlePheThrSer 375  
 Db 1066 AAAAATAATGCAAAAGATTAAGAAAGATTTAAAAAATATATATATATATATATATATAT 1125  
 OY 376 LybValIleThrAenAenAenGlyLeuAepPheAepProThrLybGlyAenAenGluIleValGly 395  
 Db 1126 ACGATTTCTTAAAAACCATCTTTATGATCCA-----AATGATTAATTTCTTGAGTGG 1176  
 OY 396 GluGlyLeuProThrPheLeuSerAenGluAepCysAlaLybLeuGlnGlnSerTyCysPhe 415  
 Db 1177 AGAAATCTGCAACATTTCTTATATGAGACCTGTGCAAGATTTGAAATCTTATGTTT 1236  
 OY 416 TybPheGluLybLybCysProAepGlyGluAenAlaCysLybAenIleArgAlaThrCys 435  
 Db 1237 TACTACGAAAAAGCTTGTCCAAATGCCAAGAAAGAGTATGAATTTGAGGCGCGGT 1296  
 OY 436 TybLybAenGlyLeuAepAlaArgAlaAenLybValIleGlnGlnLybAenMetArgGlyMet 455  
 Db 1297 TATTAAGAGGGCTTGATGACGCGGCAATTAAGTGTCTGCAGAAATATATGCGTGGTTA 1356  
 OY 456 LeuHisGlySerAenLybSerTrpLeuGluLybAepGlnGlnGlnGlnGlnValIleCys 475  
 Db 1357 TTACGTGTTCAATCAAAAGTTGCTTAAAGAGTTTCAAAAGATTTGTAAGATATG 1416  
 OY 476 GluLybLeuLybLybGluAenLybGlySerPheSerAenAepGluLeuPheIleLeuCys 495  
 Db 1417 AAGAGCTTA---AAAGAAAAATTAAGAAAGATTTCCCAACGATGAATATTTGTTGTG 1473  
 OY 496 ValGlnProAlaLybAlaAlaArgLeuLeuThrHisAepLybAenTyMetLybThrIlePhe 515  
 Db 1474 GTACAGCCAGAAAAAGCTGCAAGATTTACTTACACAGATCTCAATAGAGGCTTATCTTT 1533  
 OY 516 LeuAenGlnGlnLeuAepGlnLybLybArgAepPheProThrAepLybAenCysLybGluLeu 535  
 Db 1534 TTACGACACAAATTTGATCAAAAGAGATTTTCCGACAGATTAAGATCGCAAGGAATTA 1593  
 OY 536 GlyArgLybCysGlnAepLybGlyGluAepSerLybGluIleThrTrpProCysHisThr 555  
 Db 1594 GCGAAAAAATCCCAAGATTTTGGAAAGATTTCAAAAGAAATTTACATGCGCATGCAATACG 1653  
 OY 556 LeuGluGlnGlnCysAenAenGlnGlyThrThrGlnGluIleLeuLybGlnValIleLeuAen 575  
 Db 1654 CTGAGAGCAGCATGATCGCTTGGGACTACAGAAATTTTAAAGCAAGTTTATTTGAT 1713

[illegible]

Qy	936	AspProcyValaThrGluCysLysSerLeuGlnThrAspThrTPValaThrGlnThr	955
Db	2794	GATCGAAGGCGAAGGAAATGCAATCTTACAGCAACAGACACATGGGTTCACACACA	2853
Qy	956	SerThrAspThrSerThrSerThrIleThrSerThrIleThrSerIlysiIleThrLeuThr	975
Db	2854	TCGACACACACACACACACGTCTACTATCATCATCTCCATCATCATCAAAAATTAACATTGACA	2913
Qy	976	SerThrAspThrGlyValyLysProThrIlyeCysThrThrGlyGluGluAaPAPAlaGlyAsp	995
Db	2914	TCAACGAGGCGCATCCAAACCAACCAAGTGAACGACAGG-----GATGATGACGAAGAC	2967
Qy	996	VallybProSerGluGlyLeuArgMetSerGlyTyrPasnValMetArgGlyValIleVal	1015
Db	2968	GTGAAGCCAAATGAAGGCTTGAGGGTGAACGGCGGTGCAATGATGATGAGGGGGTATAGTA	3027
Qy	1016	AlaMetValIleSerPheMetIle	1023
Db	3028	GCAATGGTATTTTCGTTCATGATT	3051
RESULT 3			
AAZ94066			
ID	AAZ94066	standard; DNA; 3084 BP.	
XX	AAZ94066;		
AC			
XX	AAZ94066;		
XX			
DT	15-SEP-2003	(revised)	
DT	05-JUN-2000	(first entry)	
XX			
DE	Pneumocystis carinii major surface glycoprotein gene HMSG14.		
XX			
KW	Major surface glycoprotein; MSG; HMSG14; human; pneumonia; diagnosis; ds.		
XX			
OS	Pneumocystis carinii; sp. f. hominis.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..3084	
FT		/*tag= a	
FT		/product= "HMSG14"	
FT		/note= "a nucleic acid comprising residues 2839-3084 of	
FT		this sequence is specifically claimed in Claim 28"	
XX			
PN	WO200009760-A1.		
XX			
PD	24-FEB-2000.		
XX			
PP	17-AUG-1999;	99WO-US018750.	
XX			
PR	17-AUG-1998;	98US-0096805P.	
XX			
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Wei Q;		
XX			
DR	WPI; 2000-206025/18.		
XX	P-PSDB; AAY79168.		
PT	Detection of the presence of Pneumocystis carinii in specimens by		
PT	identification of major surface glycoprotein (MSG) gene sequences using		
PT	two or more oligonucleotide primers derived from human P. carinii MSG		
PT	protein encoding sequence.		
XX			
XX			
PS	Claim 27; Page 72-76; 110pp; English.		
XX			
CC	The present sequence is that of the novel Pneumocystis carinii sp. f.		
CC	hominis gene, HMSG14, which encodes a major surface glycoprotein (MSG,		
CC	see AAY79168). The gene was isolated by PCR amplification of DNA taken		
CC	from an autopsy lung sample of an HIV-infected patient with P. carinii		
CC	pneumonia. It is 1 of 7 novel, claimed genes (see AAY79165-71). The MSGs		
CC	invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs		
CC	include a highly conserved C-terminal region of approximately 100 amino		
CC	acids; this region (also claimed) corresponds to residues 2839-3084 of		

CC carini MSG-encoding genes, especially by PCR using primers directed at  
 CC the conserved region of the genes, provides a sensitive and specific  
 CC technique for the detection of P. carinii, and the diagnosis of P.  
 CC carinii pneumonia, especially in biological specimens (e.g. blood,  
 CC sputum) from immunocompromised patients such as those with HIV infection.  
 CC (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 3084 BP, 1240 A; 455 C; 676 G; 713 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 0	3084	4352, 00
Percent Similarity: 87.28%	Conservative: 80	
Best Local Similarity: 79.58%	Mismatches: 106	
Query Match: 80.43%	Indels: 26	
DB: 3	Gaps: 6	

US-10-654-416-14 (1-1023) x AA294066 (1-3084)

```

OY 1 MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAenSerTLeaSpGluGluHis 20
Db 1 ATGGCGCGGCGGTCAACGCGCAGCAAAAGGTGCACAAAATAGCATTTGATGAGAGCAT 60
OY 21 ValLeuAlaLeuLleLeuLysGluAenSpGlyLeuSerGluGluGluCysLysLysLeu 40
Db 61 GTTTAGCTTTGATTTTAAAAAATGATTGAGAAATACAAATGCAAACTTAAGTTG 120
OY 41 LysLysTyrCysGlnGluLeuThrGlnAlaLysLysAenLleGluGlnValHisArgLys 60
Db 121 GAAGAATTTTGCAGAAACATTAAACAATGCAGATTAAATCCAGAAAAAGTTCAAGAAAA 180
OY 61 LeuLysGlyPheCysGluAenSpGlyLysAlaAepThrLysCysLysGluLeuLysAlaAen 80
Db 181 TTTAAAGATTTCTTGATACCGGAAACGAAATGAAAAATGCAAGATCTTAAAAACAAA 240
OY 81 IleGluLysLysCysThrThrLleLysGlyLysLysLysGluAlaLleLysLysLysLle 100
Db 241 GTCAATCAAAATGATTAATTTTCAAGGAAAACTTCAACAGCTGCTAGAAAAAAATTT 300
OY 101 GlnLleLleThrAspLysAspCysLysGluAenGluGlnGlnCysLysPheLeuGluGly 120
Db 301 TCAGAAATTAACAGATGAGAGATTGCAAAAAAGAAATGAACAACATGCTTTTGGAGGGA 360
OY 121 ValCysSerLysGluLeuLysAspAspCysAenThrLeuArgAenLysCysTyrGlnLys 140
Db 361 GCATGTCCAAACGAACTTAAGATGACTGCATAATTAAGCAATAGCTTATCAAAAA 420
OY 141 LysArgAenLysValAlaGluGluValLeuAenArgAlaLeuAenSerAspLeuAenGly 160
Db 421 GAACGGAACATGTGCGAGAAAGATTTCTTTGAGCGCGCTTCGTGTGATCTCAATAA 480
OY 161 SerValLleCysGluLysLysLysLysLysLysLysCysProValMetGlyArgLysSerAsp 180
Db 481 ACAAGACATGTGAAAAAACTGAAAGAACTTCCGAAATTTGAAGAAAGAAAGCAAT 540
OY 181 GluLeuThrAsnLeuCysLeuAenGlnLysGluThrCysLysAsnLleLeuLleGluLys 200
Db 541 GAATTAACGAGCTTGTCTTATCAAAAAACACATGCGTAAGCTTGTAAACAAAGCA 600
OY 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaLleLeuGlySerPheLys--- 219
Db 601 AAAAGTAAATGATATCTTGAAGAAAGAGTTGAAGAACCTTAAGACAGAAATTAATG 660
OY 220 LysGluThrCysLeuGluLeuLeuGluGlnCysTyrPheTyrIleGlyAsnCysGlyAsp 239
Db 661 CGAGAAAAATGTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
OY 240 AspAsp-----IleLleLysCys 245
Db 721 GACCAATCAAAAGTCAATTAACCTAATAATAAGACCTGCAAGAAATATGATACCAAGAGT 780
OY 246 IleGluLeuGlyGlyLysCysGlnGluGlnAenLleAlaTyrMetProGlyProAsp 265

```

```

Db 781 GATGATTTAGCAAAAAGTGTGAAAAAATATTTGTTATATGATCATCCAGATCCGAT 840
OY 266 PheAspProThrArgProGluAlaThrIleAlaGluAenLleGlyLeuGluPheTyr 285
Db 841 TTGCATCAACTTAACGAGAGCTTACCTAGACAGAGACATAGGGCTTGAAGACTTAT 900
OY 286 LysLysValGluGluAenSpGlyValPheLleGlyLysAenHisLysLeuArgAspAlaThrAla 305
Db 901 AAGAGGCGAGAGAGATGAAATTTTGTGAAGAACCACTGTAAAGATGCAACAGCT 960
OY 306 LeuLeuAlaLeuLleGlnAenSerSerLeuLysLysLysAspAspLysGluLysCys 325
Db 961 TTGTGGCACTACTT-----CTTAAGAAAAACCTTAAGAAAAAGAAATGT 1005
OY 326 GluGluAlaLeuGlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluSerLeu 345
Db 1006 ATAAAGCCCTTAATAAAACCTGCGAAACCCCTCATGAACATGAGGCTTGAAGAAATCTA 1065
OY 346 CysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAenAsp 365
Db 1066 TGTAAAGAAATTAACCAAGATGATGAGAAACGAAAAATGTATGAACTAAGAAAAAGAT 1125
OY 366 IleAsnLysThrCysLysIlePheThrSerLysValThrAenAenArgLeuPheAspPro 385
Db 1126 GTTAACAAACTTTATCAAGCTTACATCAACATTTCTTAATAAACCGCTTACATTTCA 1185
OY 386 ThrLysGlyAsnAenGluLleValGlyTyrPglLysLeuProThrPheLeuSerAsnGlu 405
Db 1186 CCTGATGGA-----ATTGCGGAATGGGGAATAATTAACGACATTTCTTAGTGATA 1236
OY 406 AspCysAlaLysLeuGluSerTyrCysPheTyrPheGluLysLysCysProAspGlyGlu 425
Db 1237 GATTGTGCAAAACATGAATCTTATTTGCTTATTAATAAGAACTTGCCAGATGTCAA 1296
OY 426 AsnAlaCysLysAsnLleArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAen 445
Db 1297 GAACCTGTATGAATGTGAGGCGAGCGCTTTTCAAGAGAGGCTTGATGACGCGCAAC 1356
OY 446 LysValLeuGlnGluAenMetArgLysLysLysLysLysLysLysLysLysLysLys 465
Db 1357 AGTGTTGCAAAATAATATGCGGTATTAATGATGATTAATTAATTAATTAATTAATTA 1416
OY 466 LysPheGlnGlnGluLeuValLysValCysGluLysLysLysLysLysLysLysLys 485
Db 1417 GAGTTTCAACAAAGATTGAATAAGTATGTAAGAGCTA---AAAGAAAAATTAAGAAAT 1473
OY 486 PheSerAsnAspGluLeuPheLleLeuCysValGlnProAlaLysAlaArgLeuLeu 505
Db 1474 TTCCCAACGATGAATATTTTGTCTGTGTGATACGCCAGCAAGCTGCAGATTACTT 1533
OY 506 ThrHisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAsp 525
Db 1534 ACACACGATCATCAATAGAGGTTACCTTTTACGACAAACATTTGATGATCAAAAGAGAT 1593
OY 526 PheProThrAspLysAsnCysLysGluLeuGlnLysArgLysCysGlnAspLeuGlyGluAsp 545
Db 1594 TTTCCGAGATGAATAAGCTGCAAGAACTAGAGAAAAATCCCAAGATTGAGAAAGAT 1653
OY 546 SerLysGluLleThrTyrProCysHisThrLeuGlnGlnGlnCysAsnArgLeuGlyThr 565
Db 1654 TCAGAAAGAAATTAATGATGCGCATGTCTACCTGAGCGACGACATCGCTTGGGACT 1713
OY 566 ThrGluLleLeuLysGlnValLeuLeuAenSpGluHisLysAspThrLeuLysAspGlnGlu 585
Db 1714 ACAGAAATTTAAAGCAGGTTTATTTGATGAACACAAAGATTAATTTGAAAGACCAAGAA 1773
OY 586 SerCysValLysTyrLysLysGluLysCysAsnLysTyrPheArgArgGlyAspAspArg 605
Db 1774 AGTGTGTAAATATCTTAAGAAAGATGTAATTAATTAATTAATTAATTAATTAATTAAT 1833
OY 606 PheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValLysAspValLys 625
Db 1834 TTCTTTTGTATGTGTTTTCAGAAACGCTACGTGTGAGCTGATGTTAAAGACCTGAAA 1893

```



QY 626 AApArgCYeGIuVal1PheLYbLYeAbn1LeLYsAlaSerTYrIle1leGIuPheLeuGIu 645  
 Db 1894 GACAGGTGTGAAGTATCAAAAATAATTAAGCTTCATATATTGAATTTCTTGAA 1953  
 QY 646 AAbnAnThrAbnLYb1leThrThLeuGIuArgAbnCYbProSerTPHleThrTYrCYb 665  
 Db 1954 AATATATCAATAATAATAACAACACGTGAAGAAATGTCCTCTTGCAATACGATATGCG 2013  
 QY 666 AAbnArgPheSerProAbnCYbProGIuLeuThrLYbGIuAbnSerCYbThrLYb1leLYb 685  
 Db 2014 AATAGATTTTCACCTAATTGTCAGGCTTACGAAGAAAGATAGTGTACAAAATCAAG 2073  
 QY 686 LYbH1sArgGIuProPheTYrLYbArgLYbAlaLeuGIuAbnAlaLeuLYbValGIuLeu 705  
 Db 2074 AAGCATTTGTAGCCGCTCTAATAAAGAAAGCCCTTGAAAGATGCTCTCAAAAGTAGACCTT 2133  
 QY 706 GInGIuLYbLeuThrAbnLYbSerLYbCYbGIuProAlaLeuLYbArgTYrCYbThrVal 725  
 Db 2134 CAAGGAAATATGACTGATTAATCTAAATGTGAACCTGCATGAACAGATATTGTACAGTA 2193  
 QY 726 AAlGIuAbnValAAbnAbnAlaSer1leSerGIuLeuCYbLYbAlaAbnThrLYbAbn 745  
 Db 2194 GCGGAAACGTAAATATGCTCAATCAATGCTTATGCAAAAGCTAACACCAAGGATTAAC 2253  
 QY 746 SerGIuLYbSerAePGLuAePAlaArgLYbGIuLeuCYbGIuLYbLeuValLYbGIuVal 765  
 Db 2254 TCTCGAAAGAGTGTAGAGATGCTAGAAAGAACTCTGTGAGAAATCAGTCAAAAGAAAGTG 2313  
 QY 766 GInGIuGIuCYbLYbAlaLeuProThrGIuLeuGIuGIuProAlaAlaAePLeuLYbLYb 785  
 Db 2314 GAAGAACAGTGCACAAAGCATTTCAACAGAAATTAGACAAACGCGCAGCTGATCTAAATAA 2373  
 QY 786 AApTYrLYbThrTYrGIuLeuLYbLYbArg1leGIuGIuAlaAePLeuLYbSerSer 805  
 Db 2374 GATTATAGACATATAGAGAACTTAAGAAAGCTGACAGGAAAGCAAAAGCTCAGT 2433  
 QY 806 LeuValLeuSerLeu1leLYbLYbAeNG1uSerAbnValSerLYbSerAePLeuLYbAbn 825  
 Db 2434 CTCTTTTGTGACCTCATTAAGAAABAAACGAAGTATGATCAAAAGTAATAGCAAAAC 2493  
 QY 826 LYbAePLeuAbnAlaValSerAeNG1uLeuGIuAbnThrThLYbH1sValLYb1leLeu 845  
 Db 2494 AAGGATTAAGATGCGCTTCAACGAGCTTCAAGATACCAAAACATGTGAAATATCTA 2553  
 QY 846 ArgArgGIuValLYbAePValSerValThrGIuLeuGIuAlaLYbAlaPheAePLeuAla 865  
 Db 2554 CGGAGAGAGATTGAAGATGATCCGTAAACGAATTAGAAAGCTTAAGCATTTGATTGGCA 2613  
 QY 866 AAlGIuVal1PheGIuArgTYrVal1AsPLeuLYbGIuArgCYbAbnLYbLeuGIuSerAeP 885  
 Db 2614 GCAGAAATATTTGGAAGATATGATTTGAAGAAAGATTAATTAATTTGGAATTCGAT 2673  
 QY 886 CYbArg1leLYbGIuAePValLYbAePLeuGIuValCYbLYbLYb1leAbnLYbAla 905  
 Db 2674 TGCGAATTAAAGAGGATTGCAAAAGCTTGAAGAAAGTATCAAAAGATTAATTAAGGCT 2733  
 QY 906 CYbArgAbnLeuLYbProLeuGIuValLYbProH1sGIuThrVal1ThrGIuSerThrThr 925  
 Db 2734 TGTGCAATCTGAAGCCTCTGAGAGTGAAGCCGACCAAGAACAGTGAAGAGTACACG 2793  
 QY 926 ThrThrThrThrThrThrThrThrVal1AlaAePProLYbAlaThrGIuCYbLYbSerLeu 945  
 Db 2794 ACNACTCAACACACACACACACCGTTGCCATCCGAAGGCAAGCAATGCAATCTCTTA 2853  
 QY 946 GInThrThrAePThrTPValThrGIuThrSerThrH1sThrSerThrSerThr1leThr 965  
 Db 2854 CAGACAACAGACATGGGTATACACAGACATGACACACACAAAGCAGTCTACTATACACA 2913  
 QY 966 SerThr1leThrSerLYb1leThrLeuThrSerThrArgAGCYbLYbProThrLYbCYb 985  
 Db 2914 TCTACCATCATCAATAATAATACTTGAATCAACGAGGCGATCAAAACCAACCAAGTGT 2973

QY 986 ThrThrGIuGIuAePAPAlaGIuAePValLYbAePProSerGIuLYbLeuArgMetSer 1005  
 Db 2974 ACAGACAGG-----GATGATGCAGAAAGCCTGAAGCCAGTGAAGGCTTGAGGCTAGC 3027  
 QY 1006 G1YTPAbnValMetAAGLYbVal1leVal1AlaMetVal1leSerPheMetIle 1023  
 Db 3028 GGGTGAATGTGATGAGAGGGGGGTGATGTAGCAATGTTATTTGCTTCATGATTT 3081  
 RESULT 4  
 AA294065  
 ID AA294065 standard; DNA; 3090 BP.  
 XX  
 AC AA294065;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 05-JUN-2000 (first entry)  
 XX  
 DE Pneumocystis carinii major surface glycoprotein gene HMSG11.  
 XX  
 KM Major surface glycoprotein; MSG; HMSG11; human; pneumonia; diagnosis; de.  
 XX  
 OS Pneumocystis carinii; sp. f. hominis.  
 XX  
 FH Key  
 FT CDS  
 FT 1. .3090  
 FT /\*cag= a  
 FT /product= "HMSG11"  
 FT /note= "a nucleic acid comprising residues 2845-3090 of  
 this sequence is specifically claimed in Claim 28"  
 XX  
 PN W0200009760-A1.  
 XX  
 PD 24-FEB-2000.  
 XX  
 PF 17-AUG-1999; 99WO-US018750.  
 XX  
 PR 17-AUG-1998; 98US-0096805P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 FI Kovacs JA, Huang S, Maeur H, Fischer SH, Gill VJ, Wei Q;  
 XX  
 DR WPI; 2000-206025/18.  
 DR P-PSDB; AAY79167.  
 XX  
 PT Detection of the presence of Pneumocystis carinii in specimens by  
 PT identification of major surface glycoprotein (MSG) gene sequences using  
 PT two or more oligonucleotide primers derived from human P. carinii MSG  
 PT protein encoding sequence.  
 XX  
 PS Claim 27; Page 64-68; 110pp; English.  
 XX  
 CC The present sequence is that of the novel Pneumocystis carinii sp. f.  
 CC hominis gene, HMSG11, which encodes a major surface glycoprotein (MSG,  
 CC see AAY79167). The gene was isolated by PCR amplification of DNA taken  
 CC from an autopsy lung sample of an HIV-infected patient with P. carinii  
 CC pneumonia. It is 1 of 7 novel, claimed genes (see AA27963-69) of the  
 CC invention that encode human-P. carinii MSGs (see AA79165-71). The MSGs  
 CC include a highly conserved C-terminal region of approximately 100 amino  
 CC acids; this region (also claimed) corresponds to residues 2845-3090 of  
 CC the present sequence. Direct detection or amplification of human-P.  
 CC carinii MSG-encoding genes, especially by PCR using primers directed at  
 CC the conserved region of the genes, provides a sensitive and specific  
 CC technique for the detection of P. carinii, and the diagnosis of P.  
 CC carinii pneumonia, especially in biological specimens (e.g. blood,  
 CC sputum) from immunocompromised patients such as those with HIV infection.  
 CC (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 3090 BP; 1255 A; 447 C; 662 G; 726 T; 0 U; 0 Other;  
 Alignment Scores: 3,18e-281 Length: 3090  
 Pred. No.: 3827.00 Matches: 728  
 Score:

Percent Similarity: 81.54% Conservative: 120  
 Best Local Similarity: 70.00% Mismatches: 164  
 Query Match: 70.73% Indels: 28  
 DB: 3 Gaps: 8  
 US-10-654-416-14 (1-1023) x AA294065 (1-3090)

OY 1 MetalaatgalaValyLeaArgGlnlaAalaglyThrGlnAenSerTLeaSPGluGlnHs 20  
 DB 1 ATGGCGCGGCGGTCAACGCGCGCAAAAGGTGCAGAAATGACATTTGATGAGAGCAT 60  
 OY 21 ValLeuAlaleuLleLeuLysGluAaSPGlyLeuSerGluGlnGluCValyLysLeu 40  
 DB 61 GTTTAGCTTTGATTTAAAAAATGATTAAGAAATGACAAATGCAAACTGAATG 120  
 OY 41 LysLysTyrcyGlnGlnLeuThrcyluaLysLeuAenileGluGlnValHlsArglys 60  
 DB 121 GAAGAAATTTGCAAAACATTAAACAATGCAGATTTAAATCCAGAAAAGTTCAAGAAA 180  
 OY 61 LeuLysGlyPheCySGluAaSPGlyLysAlaAaPThrcyLysGlyLysLeuLysAlaAa 80  
 DB 181 TTTAAAGAATTTCTGTGATACCGGAAACGAATGAAAATGTCAGATCTAAAAACAAA 240  
 OY 81 IlegluLysCySvThrcThrlleLysGlyLysLeuLysGluAlaleLysLysLysle 100  
 DB 241 GTCAATCAAAATGCAATTAATTTCAAGGAAAATCTCAACAGCTGCGAAAAAAAT 300  
 OY 101 GlnleileThraSPlyAaSPCySvLysGluAaSPGluGlnGlnGlnCySvLysPheLysGlu 120  
 DB 301 TCAGAAATTAACAGATGAGGATTGCAAAAAGAAATGAACAACAATGCTATTTTGGAGGGA 360  
 OY 121 ValCySvSerLysGluLeuLysAaSPCySvAaThrcLysGlnAaLysCySvTyGlnLys 140  
 DB 361 GCATGTCCAAACGAACCTTAAGATGATGCAATTAATTAAGCAATATGCTTATCAAAA 420  
 OY 141 LysAaSPlyLysValaleGluGlnValleuLeuAaGlaLeuAaSPCySvLysAaSPGly 160  
 DB 421 GAACGAAACAATGTGCGAAGAAAGTTCTTTGAGGCGCTTCGCGATCTCAATGA 480  
 OY 161 SerValilecyGluLysLeuLysGluLysGluLysCySvProvalMetGlyAaSPGlyAaSP 180  
 DB 481 ACAAGAATGTAAGAAAAAGCTGAAGAAGTTTCCGAAATTTGAAGAAGAAAGCAT 540  
 OY 181 GlnLeuThraSPlyAaSPCySvLysAaSPGlnLysGlnLysGlnLysGlnLysGlnLys 200  
 DB 541 GAATTAACGAGCTTTGTCTTATCAAAAAACAATGCTTAAGCTTGTAAACAAAAGA 600  
 OY 201 AaSPlyLysCySvGlyThrcLysLysThraSPValSerAlaleuGlySerPheLys--- 219  
 DB 601 AAAAGTAATGATATCTTTGAAAAAGAAAGTTGAAGAAGCATTAAGAAAGAAATTTG 660  
 OY 220 LysGluThrcyLysLeuGlnLeuLysGlnLysCySvTyrcPheTyrlleGlyAaSPGlyAaSP 239  
 DB 661 CGAGAAAATGTCTATCTTACTTGAACAATGTTCTTCAAGAGGGAACGTGAAAGA 720  
 OY 240 AaSPaSP-----IleleLysCyS 245  
 DB 721 GACCAATCAAAAGTGCATTAACCTAATTAATAAGACCTGAAGAATATGATCAAGAGTGT 780  
 OY 246 IlegluLeuGlyLysLysCySglnGlnGlnAenileAlaTyrcMetProSPGlyProAaSP 265  
 DB 781 GATGAATTAAGCAAAAAGGTGGAAGAAGAAATATGTTATATCATCTCCAGAAATCCAT 840  
 OY 266 PheAaSPProThraSPGlnLysAlaThrcLysAlaGlnAaSPGlyLysGlnGlnLysPheTyrc 285  
 DB 841 TTTCATTCACAATCAACCAAGCCTTACATGACAGAGGACATGAGCTGGAAGAAGCTTTAT 900  
 OY 286 LysLysValGlnGlnAaSPGlyValPheileGlyLysAaSPHlsLeuAaSPAlaThrcAla 305  
 DB 901 AAGAAGGCGGAAGAGATGAAATTTTGTGGAGAACAAACATGTAAGAGATCAACAGCT 960  
 OY 306 LeuLeuAlaleuLleileGlnAaSPSerLeuLysLysLysAaSPlyGluLysCyS 325

DB 961 TTGTGGCACTACTT-----CTTAAGAAAAACCTTAAGAAAAAGAAATCT 1005  
 OY 326 GlnGlnAlaleuGlnLysSerCySvLysAaSPProHlsGlnHlsGlnAlaleuGlnSerLeu 345  
 DB 1006 ATAAAGCCCTTAATAAAAACTGCGAAAACCCCTCATGAACATGAGCGCTTGAAGAACTTA 1065  
 OY 346 CySvLysLysAaSPGlyLeuSerAaSPGlyThrcLysCySvGlnGlnLysGlnAaSP 365  
 DB 1066 TGTAAAGAAATTAACCAAGATGATGAAACGAAAAATGTGATGAACTAAGAAAAAGAT 1125  
 OY 366 IleAaSPThrcyLysIlePheThrcSerLysValThraAaSPAlaGlyLeuPheAaSPPro 385  
 DB 1126 GTTAACAACCACTTACAAAGCTTACATCAACATCTTAATAACCGCTTTACATTTCA 1185  
 OY 386 ThrcLysGlnAaSPGlnLysAlaleGlyTyrcPheLysLeuProThraPheLeuSerAaSP 405  
 DB 1186 CCTGATGA-----ATGCGGAATGGGGAATAATTAACGACATTTCTTATGATGATA 1236  
 OY 406 AaSPCySvAlaleuGlnLysSerTyrcCySvPheTyrcPheGlnLysCySvProAaSPGly 425  
 DB 1237 GATTGTCAAAACCTGAATCTTATGCTTTATTAAGAAAGCTTGTCCAGATGTCAA 1296  
 OY 426 AaSPAlaCySvAaSPHlsAaSPAlaThrcCySvTyrcLysAaSPGlyLeuAaSPAlaAaSP 445  
 DB 1297 GAACCTTGTATGAATGTGAGGCGACCGCTTTATTAAGAGAGGCTTGATGCAAGGCAAAAC 1356  
 OY 446 LysValleuGlnGlnAaSPMetArgGlyMetLeuHlsGlySerAaSPLeuSerTyrcPheLys 465  
 DB 1357 AGTGTGTGCAAAAATATATGCGAGGGTATTTGATGAGCTCAAAATTAAGATTTGCTTAA 1416  
 OY 466 LysPheGlnGlnLysLeuValLysValCySvGlnLysLeuLysGlnAaSPGlySer 485  
 DB 1417 AAATTTCAACAAAGATTAAGCAAAAAGTATGTGAAGAACTG--AAAGAAATTAAGGAAGT 1473  
 OY 486 PheSerAaSPGlyLeuPheileLeuCySvAlaGlnProAlaLysAlaAaSPLeuLeu 505  
 DB 1474 TTCTGCAACATGAATGTTGTTCTGTGTATACACAGCAAGAGCGACGATTA 1533  
 OY 506 ThrcLysPheLysAaSPGlyThrcLysPheLysAaSPGlnGlnLysAaSPGlnLysAaSP 525  
 DB 1534 ACACATCAACCAATTAAGATGAGTTATCTTTTGAACAACAACATGATCAAAAGAGAGAT 1593  
 OY 526 PheProThraSPlyAaSPCySvLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 545  
 DB 1594 TTTCCGACAGATTAAGACTGCAAGAAATTAAGGAAAAATGCAAGATTTAGGAAGAGAT 1653  
 OY 546 SerLysGlnLysThrcProCySvHlsThrcLysGlnGlnGlnCySvAaSPGlyLys 565  
 DB 1654 TCAAAAGAAATTAACATGCGCAATGCTATCACTAACAACGCAATGCAATCGCTTAGGAGAT 1713  
 OY 566 ThrcLysLeuLysGlnValleuLeuAaSPGlnHlsLysAaSPThrcLysLysAaSPGlnLys 585  
 DB 1714 ACAGAAATTTTAACAAGATTTTATGGAAGAACACAACAAGATTAATTAAGAAAGTCAATA 1773  
 OY 586 SerCySvAlaleGlyLysLysGlnLysCySvAaSPHlsTyrcPheTyrcGlyLysAaSPGly 605  
 DB 1774 AACGTGCAAAATTTTAAGAAAGAAATGCAATTAAGTGAAGAGGAGTGAAGATGAT 1833  
 OY 606 PheSerPheValCySvAlaPheGlnAaSPAlaThrcCySvGlnLysMetValLysAaSPValLys 625  
 DB 1834 TTTTCTTTGATGTGTTTTCAAAACGCTTCACTGACGTGATGCTTAAGAGAGCTGCAA 1893  
 OY 626 AaSPGlyGlnLysAlaPheLysLysAaSPHlsLysAlaSerTyrlleileGlnPheLysGln 645  
 DB 1894 GATAGGTGCAAAATTAATGGAAGAAATTAAGCAACATGATTAATTAAGATTCCTTA 1953  
 OY 646 AaSPHlsThraSPlyLysThrcLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 665  
 DB 1954 AAAAATCAAAATTAAGCAAGATCAAGCAAAATTTGTCTCCCTGATGCAATCAACATG 2013  
 OY 666 AaSPHlsPheSerProAaSPGlyLeuThrcLys---GlnAaSPCySvThrcLysIle 684  
 DB 2014 GATGATTTTAACCAATTTGCTGATCTTAAGAAAGAAAGAACTTTGTGCAAAATCTT 2073

```

QY 685 LysLeuHisArgGluProPheTyrLeuArgValLeuGluAAspAlaLeuLysValGlu 704
Db 2074 AAAAAAAAAATTCGCAACCATCTTACAAAAGAAAGTTTAAAGATGCTTTAAAGTGAAG 2133
QY 705 LeuGlnGlyLeuLeuThrAspLysSerLysCysGluProAlaLeuLysArgTyrCysThr 724
Db 2134 CTTTCAGGAATTAATTAAGATATATACTAAATGTGAACCTGCATAGAAAGATATTGTGACA 2193
QY 725 ValAlaLeuVenValaAAsnAlaSerLysLeuGlyLeuGlyValAlaAsnThrLysAsp 744
Db 2194 GTATTGAAAGCGTAAATTAATGCTCATACGAGCTTATGTAAAGATTAATACCGAAAGT 2253
QY 745 AAsnSerGlyLysSerAsp---GluAAspAlaArgLysGluLeuGlyLysLeuValLys 763
Db 2254 AAAAACTAAAAAGCCGATTAATAAAATGTTAGAAAGAAAGCTTTCTTAAATTTGTGAA 2313
QY 764 GluValGluGlnGlnCysValAlaLeuProThrGluLeuGlnProAlaAlaAspLeu 783
Db 2314 GAGGTGGAACGCAATGCAAGATATTACCAACAGAAATTAAACAGAGCTGAAAAAAGTCTA 2373
QY 784 LysLeuAspTyrLysThrTyrGluGluLeuLysLysArgAlaGluGluAlaMetAsnLys 803
Db 2374 AAAAAAGATGTTAAGACATATAGAGAACTTAAAGAAAGCGCAAAAAAAGCAATGAAACAG 2433
QY 804 SerSerLeuValLeuSerLeuLysLeuLysAsnGluSerAsnValSerLysSerAsnSer 823
Db 2434 TCCAGCCTTGTATTATCACTTTAGAAAAACAAAGTAAATACATGAAAAATTAATAGC 2493
QY 824 LysAsnLysAspLysAAsnAlaValSerAsnGlyLeuGlnAspThrThrLysHisValLys 843
Db 2494 AAAAAACAAGTAAAGATATGCTTTCAACGCACTTCAAGTACCACAAAATATGTGAAA 2553
QY 844 IleLeuArgArgGlyValLysAspValSerValThrGluLeuGluAlaLysAlaPheAsp 863
Db 2554 ATACTAGAAAGAGAGCTTAAAGAGCACTTGTAACAGAAATGGAAGCCAAAGCACTTTGAT 2613
QY 864 LeuAlaAlaGluValPheGlyArgTyrValaAspLeuLysGluAsnLysLeuGlu 883
Db 2614 TTGGCAGCAAGAGCTTTGAAAGATATGTAGACTTTGAAAGAAATGTGAAATTTGACT 2673
QY 884 SerAspCysArgLysLysGluAspCysLysAspLeuGlnGluValCysLysValAsn 903
Db 2674 TCGGATTGCGGAGTTAAAGACATTTGGAGTGTAAAGAAAGTGTGGAAAGATTGAG 2733
QY 904 LysAlaCysArgAsnLeuLysProLeuGlnValLysProHisGluThrValThrGluSer 923
Db 2734 AAGACATGTCAGCATCTGAAGCTCTGAGAGTGAAGTCGCAATGAATATGTCACAGAAAGC 2793
QY 924 ThrThrThrThrThrThrThrThrThrThrValAlaAspProLysValaThrGluCysLys 943
Db 2794 ACAACGACGACACACACACAAACACACCTTACCCGATCCGAAAGGCAACAAATGCAAA 2853
QY 944 SerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHisSerThrSerThr 963
Db 2854 TCCTTACGAGCAACAGATACATGGTTTACACAGACATGACACACACAGACAGCTTACC 2913
QY 964 IleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArgCysLysProThr 983
Db 2914 ATCATCATCTACCATACATCAAAAATACATTCATCAACAGAGCGCATGCAACCAACC 2973
QY 984 LysCysThrThrThrGlyGluGluAspAlaGlyAspValLysProSerGluGlyLeuArg 1003
Db 2974 AAGGTATCGACAGGG-----GATGACACAGAGAGCGTGAACCGAGTGAAGATTGAAG 3027
QY 1004 MetSerGlyTyrAsnValaMetArgGlyValIleValaMetValIleSerPheMetIle 1023
Db 3028 ATGATGGGTGAGACGCGATGAGAGGGGGGTGATAGCAATGTTATTCCTTCATGATT 3087

```

RESULT 5  
AAZ94067  
ID AAZ94067 standard; DNA; 3081 BP.

```

AC AAZ94067;
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE Pneumocystis carinii major surface glycoprotein gene HMSG32.
KM Major surface glycoprotein; MSG; HMSG32; human; pneumonia; diagnosis; ds.
XX
OS Pneumocystis carinii; sp. f. hominis.
XX
FH Key
FH CDS
FT 1. 3030
FT /*tag= a
FT /product= "HMSG32"
FT /note= "a nucleic acid comprising residues 2836-3081 of
FT this sequence is specifically claimed in Claim 28"
XX
XX MO200009760-A1.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99MO-US018750.
XX
XX 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VI, Mei Q,
XX WPI: 2000-206025/18.
XX DR P-PSDB; AAY79169.
XX
XX
XX Detection of the presence of Pneumocystis carinii in specimens by
XX PT identification of major surface glycoprotein (MSG) gene sequences using
XX PT two or more oligonucleotide primers derived from human P. carinii MSG
XX PT protein encoding sequence.
XX
XX
XX Claim 27; page 80-84; 110pp; English.
XX
XX The present sequence is that of the novel Pneumocystis carinii sp. f.
XX CC hominis gene, HMSG32, which encodes a major surface glycoprotein (MSG,
XX CC see AAY79169). The gene was isolated by PCR amplification of DNA taken
XX CC from an autopsy lung sample of an HIV-infected patient with P. carinii
XX CC pneumonia. It is 1 of 7 novel, claimed genes (see AA79165-71). The MSGs
XX CC invention that encode human-P. carinii MSGs (see AA79165-71). The MSGs
XX CC include a highly conserved C-terminal region of approximately 100 amino
XX CC acids; this region (also claimed) corresponds to residues 2836-3081 of
XX CC the present sequence. Direct detection or amplification of human-P.
XX CC carinii MSG-encoding genes, especially by PCR using primers directed at
XX CC the conserved region of the genes, provides a sensitive and specific
XX CC technique for the detection of P. carinii, and the diagnosis of P.
XX CC carinii pneumonia, especially in biological specimens (e.g. blood,
XX CC sputum) from immunocompromised patients such as those with HIV infection.
XX CC (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 3081 BP; 1231 A; 431 C; 663 G; 756 T; 0 U; 0 Other;

```

Alignment Scores:

Pred. No.:	Length:	Score:
7.57e-249	3081	3401.50
Percent Similarity:	76.17%	Matches: 655
Best local Similarity:	62.68%	Conservative: 141
Query Match:	62.86%	Mismatches: 208
DB:	3	Indels: 41
		Gaps: 12

US-10-654-416-14 (1-1023) x AAZ94067 (1-3081)

QY 1 MetAlaArgAlaValLysArgGluAlaAlaGlyThrGlnAsnSer-----IleAspGlu 18

Db 1 ATGGCG 30

QY 19 GluHisValLeuAlaLeuAlaLeuLysGluAspGlyLeuSerGluGlnGlnCysLysLys 38



```

QY 739 LysAlaenThrLysAspSerGlySerSerAspGluAspAlaArgLysGluLeuCys 758
DB 2230 AAGATTAATACCGAAGAACTAAAGAAAGCCGATTAATGAAGTAAAGAAAGACCTTGT 2289
QY 759 GluLysLeuValLysGluValGluGluGlnCysLysAlaLeuProThrGluLeuGln 778
DB 2290 CTAAATTAAGTGAAGAGGTGAACAGCAATGTAATAAGTAAACAGCAAGATTGACAT 2349
QY 779 ProAlaAlaAspLeuLysAspTyrLysThrTyrGluGluLeuLysLysArgAlaGlu 798
DB 2350 GAGGAAAAAGCCTAAAGATGATTTGAAACATTTGAAAACTTAAAAAAACAGCAGAG 2409
QY 799 GluAlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGluSerAsnVal 818
DB 2410 AAAACAATGAATTAATCCATCTTTTATCAATTCCTTAAGAAAGATGAATAATATCA 2469
QY 819 SerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGlnAspThr 838
DB 2470 TCGAAAAATAGTACAAAGACAGATAGAAATACCGTTCAACGAGCTTCAAGATACC 2529
QY 839 ThrLysIleValLysLysLeuArgGlyValLysAspValSerValThrGluLeuGlu 858
DB 2530 ACAGAACATATGAATAATCTACGAGAGAGATTAGAGATGATCCGTAACAGAACTGAA 2589
QY 859 AlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTyrValAspLeuLysGluArg 878
DB 2590 GCTAAGGATTTGATTTGTCGACAGAAATATTGGAAGATATCTAGACTTAAGAAAGAA 2649
QY 879 CysAsnLysLeuGlnSerAspCysArgLysGluAspCysLysAspLeuGluVal 898
DB 2650 TGTAAATTAATGGAATCAGATTTGCGAGAGTTAAGGAGATTCGAAAGATTTGAAGAGATA 2709
QY 899 CysLysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLysProHisGlu 918
DB 2710 TGTGAAAGATACCAAGAGATGTTCCGAAATTAACCACTGAAGAGTAAACCCGACAGAA 2769
QY 919 ThrValThrGlnSerThrThrThrThrThrThrThrThrThrThrValAlaAspProLys 938
DB 2770 ACAGTACAGAAAGCAACAGACAGCCGACGACACAGACCGTTACTGATCCGAAG 2829
QY 939 AlaThrGluCysLysSerLeuGlnThrThrAspThrThrValThrGlnThrSerThrHis 958
DB 2830 GCAACAGATGCAATCTTTACAGCAACAGATATCATGATTAACAGACTTCGACAT 2889
QY 959 ThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArg 978
DB 2890 ACCAGCAGCTACATCATCATCAATCAATCAAAAAAATACATCAATCAATCAACAGG 2949
QY 979 ArgCysLysProThrLysCysThrThrGluGluGluAspAlaGlyAspAspLysPro 998
DB 2950 CGTTGCAAAACCAACAGTGAACAGACAGG-----GATGATGCGAGAGAGAGTGAAGCG 3003
QY 999 SerGluGlyLeuArgMetSerGlyTyrAsnValMetArgGlyValIleValAlaMetVal 1018
DB 3004 AGTAGGAGATTGAAGATAGTGGTGAACATGATGAGGGGGTGAATGTAAGCAATGGTT 3063
QY 1019 IleSerPheMetIle 1023
DB 3064 ATTTCGTTCATGATT 3078

```

## RESULT 6

AA294063 standard; DNA; 3042 BP.

```

XX AA294063;
AC AA294063;
XX 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX Pneumocystis carinii major surface glycoprotein gene HMSGp1.
XX Major surface glycoprotein; MSG; HMSGp1; human; pneumonia; diagnosis; ds.
KW

```

```

XX OS Pneumocystis carinii; sp. f. hominis.
XX FH Location/Qualifiers
FT CDS
FT 1..3042
FT /tag= a
FT /product= "HMSGp1"
FT /note= "a nucleic acid comprising residues 2894-3042 of
FT this sequence is specifically claimed in Claim 28"
XX
XX W0200009760-A1.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99WC-US018750.
XX
XX 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
XX WPI; 2000-206025/18.
XX P-PSDB; AAY79165.
XX
XX Detection of the presence of Pneumocystis carinii in specimens by
XX identification of major surface glycoprotein (MSG) gene sequences using
XX two or more oligonucleotide primers derived from human P. carinii MSG
XX protein encoding sequence.
XX
XX Claim 27; Page 49-53; 110pp; English.
XX
XX The present sequence is that of the novel Pneumocystis carinii sp. f.
XX hominis gene, HMSGp1, which encodes a major surface glycoprotein (MSG)
XX see AAY79165). The gene is 1 of 7 novel, claimed genes (see AA279063-69)
XX of the invention that encode human-P. carinii MSGs (see AAY79165-71). The
XX MSGs include a highly conserved C-terminal region of approximately 100
XX amino acids; this region (also claimed) corresponds to residues 2894-3042
XX of the present sequence. Direct detection or amplification of human-P.
XX carinii MSG-encoding genes, especially by PCR using primers directed at
XX the conserved region of the genes, provides a sensitive and specific
XX technique for the detection of P. carinii, and the diagnosis of P.
XX carinii pneumonia, especially in biological specimens (e.g. blood,
XX sputum) from immunocompromised patients such as those with HIV infection.
XX (Updated on 15-SEP-2003 to standardise OS field)
XX
XX Sequence 3042 BP; 1307 A; 381 C; 606 G; 748 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 3,13e-200 Length: 3042
XX Score: 2762.50 Matches: 541
XX Percent Similarity: 69.10% Conservative: 177
XX Best Local Similarity: 52.07% Mismatch: 280
XX Query Match: 51.05% Indels: 41
XX DB: 3 Gaps: 18
XX
XX US-10-654-416-14 (1-1023) x AA294063 (1-3042)
QY 1 MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluGluHis 20
DB 1 GTGGCGCGGGGCTTAAGCGGACAGGTAAACAGACCA---TCAGGAGTGAATGAAGAGAA 57
QY 21 Val-----IleuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLys 38
DB 58 GTGGCTCTTTGGCTTTAATCTAAAGAAAGATTCTAAGATGATGAATAAATGCGAAGAA 117
QY 39 LysLeuLysLysTyrCysGlnGluLeuThrGluAlaLysLeuAsnIleGluGlnValHis 58
DB 118 AATTAAGAAACCTTCGAAAGATTAAGTGAACCAATCTAATCCAGAAACAACTCAT 177
QY 59 ArgLysLeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLys 78
DB 178 GAAGAATTAAAGATTTCTGTGATAGCAAAAAACGTGATTAATAAATGTAAGAACTAAAA 237

```













```
Db 3053 GAAACAAAGCTGATGCGAAGCTGATGAAAGACCGTTGAGAGACTGTACAGAAACC 3112
Qy 941 -----GluCysLeuGlnThrAspThr 950
Db 3113 AAGCTAGTAGTGAGAGAAAGTAACAGAGAGTCAATGATCAAAACAGATGACA 3172
Qy 951 TrpValThrGlnThrSerThrIleThrSerThrIleThrSerThrIleThrSer 970
Db 3173 TGGGTGACGACTGACTGCTTTCATACAGATGCGACACAGAGTACCGTACCGTACG 3232
Qy 971 LysIleThrLeuThrSerThrIleThrSerThrIleThrSerThrIleThrSer 988
Db 3233 AAGGTGACGCTGACTGCTGATCGAGAGTCAAGCTCAATGATGACACCGATTCAGC 3292
Qy 989 -----GluGluAspAspAlaGlyAspValLysProSerGluGlyLeuArgMet 1004
Db 3293 AAAGAGACACAGAAAGAAAGATGATGAGAGTGAACCGAATGAGAGGAATGAAATA 3352
Qy 1005 SerGlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db 3353 AGAGTCTCTGATGATGATTAATAATGTTGCGAGATGATTGTATGGGATGATG 3409

RESULT 9
AAQ41230
ID AAQ41230 standard; DNA; 3521 BP.
XX
AC AAQ41230;
XX
DT 27-AUG-2003 (revised)
DT 17-DEC-2001 (revised)
DT 02-SEP-1993 (first entry)
XX
DE Gene encoding major surface GP of rat P. carinii.
KW Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;
KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.
XX
OS Pneumocystis carinii.
XX
FH Key Location/Qualifiers
FT misc_feature 1..722
FT /*tag= b
FT /note= "fragment determined by PCR"
FT CDS 146..3412
FT /*tag= a
FT misc_feature 626..3521
FT /*tag= c
FT /note= "fragment from original GP3 clone"
XX
PN USN7958683-N.
XX
PD 01-APR-1993.
XX
PF 09-OCT-1992; 92US-00958683.
XX
PR 09-OCT-1992; 92US-00958683.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Kovace JA, Angus CW, Powell F, Edman JC;
DR WPI: 1993-159487/19.
XX
PI P-PSDB; AAR36706.
XX
XX Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
XX copy gene family - used in a vaccine and diagnostic assay for P. carinii
XX infection in AIDS patients.
XX
XX Disclosure, Page 41; 50pp, English.
XX
XX Multiple clones were identified by immunoscreening a rat P. carinii cDNA
XX library using rat serum generated against whole rat P. carinii. Clones
```

```
CC reactive with polyclonal serum were evaluated to identify those encoding
CC gp16. Three such clones (PC3, PC5 and PC14) were sequenced and contained
CC ORFs encoding closely related but distinct proteins. Although none of the
CC clones contained the complete coding sequence, overlapping regions
CC allowed alignment of the three clones and generation of a putative
CC composite sequence encoding a protein of ca. 122 kD. PCR was utilized to
CC determine the full sequence of the gene. The 5' end of the message was
CC identified by anchored PCR using primer JK58, which spanned the putative
CC start codon of the composite protein. The intervening region was
CC determined by reverse transcription followed by PCR using primers
CC spanning the 5' end to base 722 in GP3. A single clone was identified
CC that had an identical sequence to the first 76 bases of GP3. The
CC complete, composite cDNA contained an ORF encoding a protein of ca. 122
CC kD. The gp16 of P. carinii can be used in a vaccine against infection in
CC HIV-infected individuals and also as a diagnostic agent. NB. Due to a
CC poorly reproduced sequence several unreadable bases are represented by N
CC in the sequence below. See also AAQ41223-39. (Note: Revised entry
CC submitted to correct the patent number format of US Government-owned NTIS
CC applications to prevent clashes with ongoing US granted patent numbers.
CC For further information please visit the Derwent web site at
CC www.derwent.com/dwpl/updates/ntis_us.html.) (Updated on 27-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 3521 BP; 1238 A; 385 C; 695 G; 725 T; 0 U; 478 Other;

Alignment Scores:
Pred. No.: 2,06e-70 Length: 3521
Score: 1057.50 Matches: 333
Percent Similarity: 43.21% Conservative: 151
Best Local Similarity: 29.73% Mismatches: 469
Query Match: 19.54% Indels: 170
DB: 2 Gaps: 32

US-10-654-416-14 (1-1023) x AAQ41230 (1-3521)
Qy 1 MetAlaArgAlaValLysArgGlnAlaIle--GlyThrGlnSerIleAspGluGlu 19
Db 146 ATGGCAGCGCGGTTAAGAGCAAGCAGTACAAGACACACAGATGATGATGAGAAA 205
Qy 20 HisValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLys 39
Db 206 CANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGAACAAACAAATCCAAAGAA 265
Qy 40 LeuLysLysTyrCysGlnGluLeuThrGluAlaLysLeuAsnIleGluGlnValHisArg 59
Db 266 CTCGAGAAATATTGTAAGAGTTGAAGAGACACATAAATGTAGCAATGTGATGAT 325
Qy 60 Lys-LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAl 79
Db 326 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAAAAAATGCAGACGCGAAAA 385
Qy 79 AsnIleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLys 99
Db 386 AAAAGTTGAAGATGATTAATA-----GATTTTGAAGAGCAACTTCAAAAAGT 433
Qy 99 sIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnCysLeuPheLeuG 119
Db 434 ATTGAATTAATTAATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTACTTTTAA 492
Qy 119 uGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrG 139
Db 493 AGAGAGCGATTAGATGTTTATTAAGATTAAGTATGATGATGAGGAGAGATTTCAA 552
Qy 139 nLysLysArgAspLysValAlaGluGluValIleLeuLysArgAlaLeuArgSerAspLeu 159
Db 553 ATTGAACCGTCAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGGGATGCTAA 612
Qy 159 nGlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGlu 179
Db 613 AGAAGAACTTAATGTAAGAGAAAGATGATCTGTTGCCAGTGTGAGCCGAGAAAG 672
Qy 179 rAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleG 199
```





QY 839 ThrLeuHisValIys-----IleLeuArgGlyValIysAspValSerValThrGlu 856  
 DB 2759 NNGAAGATACAAACCTTGACCTGTTAAAGACGATGTAGATGAGGTGATACGAA 2818  
 QY 857 LeuGluAlaIysAlaPheAspLeuAlaGluValPheGlyArgIleValAspLeuLys 876  
 DB 2819 GTAGAGGTCAAAAGCATTTGATGCAACGACGATACCATTTGCAATTGTTGGAAATTGAAA 2878  
 QY 877 GluArgCysAsnLysLeuGluSerAspCysArgIleIleValAspGlyValAspLeuGlu 896  
 DB 2879 GAAAGATGTAAAGCTTTAGAAATTGATTCGCTTTTAAAGAGATGTGCCAGATCTAA 2938  
 QY 897 GluValCysArgLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLysPro 916  
 DB 2939 CAAGCTTGCAGAAATATAGACACTTATGTAAA---CTGGAAACATTTAGAAATTAAGNN 2995  
 QY 917 His-----GluThrValThrGluSerThrThrThrThrThrThrThr 930  
 DB 2996 NANNATACAGAGAAATATACAGAAACAAAGCGAAACGAAACGAAACGAAACGAA 3055  
 QY 931 ThrThrThr-----ValAlaAspProLysAlaThr----- 940  
 DB 3056 ACAAAAGCTAGTGCAGAGGCTGATGAAAGACCGCTTGAGAAAGCTGTTACAGAAANNAN 3115  
 QY 941 -----GluCysLysSerLeuGlnThrThrThrThrThrThr 951  
 DB 3116 NNNNNAAGTGGAGAAAGTAAACAGAAAGTGTACATGTACAAACACATACATG 3175  
 QY 952 ValThrGlnThrSerThrIleThrSerThrSerThrIleThrSerThrIleThrSerLys 971  
 DB 3176 GTGACGAGTACGTCATTTGACATACGACGACGACGAGTACGTCACGANNNNNNNN 3235  
 QY 972 IleThrLeuThrSerThrArgArgCysValAspProThrLysCysThrThrGly----- 988  
 DB 3236 NNNNNNGTTCCTTGATGCGAAGTGCAGAGCTTACCAAAATGTATACACGATTCAGCAAA 3295  
 QY 989 -----GluGluAspAspAlaGlyAspValLysProSerGluGluLeu 1002  
 DB 3296 AGAGACACAGAAAGAGATGATGATGAAGAAGTAAACCGAATGAGGAATG 3347  
 RESULT 10  
 AAQ41225 ID AAQ41225 standard; DNA; 2190 BP.  
 XX  
 AC AAQ41225;  
 XX  
 DT 27-AUG-2003 (revised)  
 DT 17-DEC-2001 (revised)  
 DT 02-SEP-1993 (first entry)  
 XX  
 DE Clone PCI4 encoding major surface gp of rat P. carinii.  
 XX  
 KW Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;  
 KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.  
 XX  
 OS Pneumocystis carinii.  
 XX  
 PN USN7958683-N.  
 XX  
 PD 01-APR-1993.  
 XX  
 PF 09-OCT-1992; 92US-00958683.  
 XX  
 PR 09-OCT-1992; 92US-00958683.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Kovace JA, Angus CM, Powell F, Edman JC;  
 XX  
 XX WPI; 1993-159487/19.  
 XX  
 PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-copy gene family - used in a vaccine and diagnostic assay for P. carinii

PT infection in AIDS patients.  
 XX  
 PS Disclosure; Page 36; 50pp; English.  
 XX  
 CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA  
 CC library using rat serum generated against whole rat P. carinii. Clones  
 CC reactive with polyclonal serum were evaluated to identify those encoding  
 CC gp116. Three such clones (PC3, PC5 and PCI4) were sequenced and contained  
 CC ORFs encoding closely related but distinct proteins. Although none of the  
 CC clones contained the complete coding sequence, overlapping regions  
 CC allowed alignment of the three clones and generation of a putative  
 CC composite sequence encoding a protein of ca. 122 kD. The gp116 of P.  
 CC carinii can be used in a vaccine against infection in HIV-infected  
 CC individuals and also as a diagnostic agent. Study of the expression of  
 CC the protein should lead to understanding its role in the pathogenesis of  
 CC P. carinii pneumonia and may lead to new strategies for control or  
 CC prevention of infection. The DNA sequence may be used to make PCR primers  
 CC for diagnostic use. See also AAQ41223-39. (Note: Revised entry submitted  
 CC to correct the patent number format of US Government-owned NTIS  
 CC applications to prevent clashes with ongoing US granted patent numbers.  
 CC For further information please visit the Derwent web site at  
 CC www.derwent.com/dmpi/updates/ntis\_us.html.) (updated on 27-AUG-2003 to  
 CC correct OS field.)  
 CC  
 SQ  
 Alignment Scores:  
 Pred. No.: 2.98e-70 Length: 2190  
 Score: 1052.00 Matches: 241  
 Percent Similarity: 51.39% Conservative: 128  
 Best Local Similarity: 33.57% Mismatches: 239  
 Query Match: 19.44% Indels: 110  
 DB: 2 Gaps: 17  
 US-10-654-416-14 (1-1023) x AAQ41225 (1-2190)  
 QY 391 GluIleValAlaGlyTrpGluGlyLeuProThrPheLeuSerAsnGluAspCysAlaIysLeu 410  
 DB 7 AAACCTTTATCGTGGGAGACGCTTCCAAACCTTTTATTAAGAGAGCTGTGCAACTT 66  
 QY 411 GluSerTrpCysPheTrpPheGluLysLysCysProAsp-----GlyIleAsnAlaCys 428  
 DB 67 GAGTCGGAATGCTTCTATTTTAAAGAAATGCCGTGTACGAATAGATTGGTGAA---GCAATGT 123  
 QY 429 LysAsnIleArgAlaThrCysTrpLysArgGlyLeuAspAlaArgAlaAsnLysValLeu 448  
 DB 124 CAAATGTACGATCAGCTGCTATTAAGGACGACAAACGATGTTGAAATACCTGTTT 183  
 QY 449 GlnGluAsnMetArgGlyMetLysIleGlySerAsnLysSerTrpLeu-----Glu 465  
 DB 184 CGAGAGAGATGAAGGAAAGCTT-----GCTAATATTAATAATTTTAATGATACGAA 237  
 QY 466 LysPheGlnGlnGluLeuValLysValCysGluLysLeuLysGluAsnLysGlySer 485  
 DB 238 AGTTGCAAAAATACAGTGCAGAAAAAGTGCAGAACTTGATAAAGATACCTTCAAAA 297  
 QY 486 PheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAlaAlaArgLeuLeu 505  
 DB 298 -----TGTCTTACCTCAAAAGACTATGTATGTCTT 330  
 QY 506 ThrHisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAsp 525  
 DB 331 TCAGATGATATTTTCTTCAATCAAAAGACTTAAAGTGTCTTTTAAATGATCAGAGAT 390  
 QY 526 PheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGlnAspLeuGlyLysAsp 545  
 DB 391 TTCATTAAGAAAAGATGTGTGTAATGGGAGAGAGATGATGAACCTTGATAGAT 450  
 QY 546 SerLysGluIleThrTrpProCysHisThrLeuGlnGlnGlnCysAsnArgLeuGlyThr 565  
 DB 451 TCATTTATTAATTAAGAAAAGTGTATTAACATTAAGAAAGAGCTGTGAATACCTTAAAGCTT 510  
 QY 566 ThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeuLysAspGlnGlu 585

















Db 529 ACAAAAGATGAATGTTTATTTAAGTGGCAACATGAGATGAAAAATTGAGAA 588  
Qy 644 LeuGluuAenAenThrAenLysIleThrThrLeuGluAenGlyProSerThrIleThr 663  
Db 589 ATTGGAAGAAAAAGCAATGAAACAGCACTCGAAGAACTCTGCACACATGGGGCCGA 648  
Qy 664 TyrCyAenAenGProSerProAenCyAPro-----GlyLeu 675  
Db 649 CATTGTACACCACTTATGAGAAATTGTCCGATGACCTTGAAGAAAAAGAAATGGCAAT 708  
Qy 676 ThrLysGluAenSerCyThrLysIleLysLysIleSarGluProPheTyrLysAryLys 695  
Db 709 GACAAATATCATATGACTGGAAGAACTGCAATGAAAAATGCACTGATACCTTAAAGATGG 768  
Qy 696 AlLeuGluAenPalAleuLysValGluLeuGlnGlyLysLeuThrAapLysSerLysCys 725  
Db 769 AAATTAAAGGAGGAGCTGACTCATCTGTTGAAAGCAGCTTAAAGATTAAGATGAATGT 828  
Qy 716 GluProAlaLeuLysAryTyrCyThr---ValAlaGlyAenValAenAenAlaSerIle 734  
Db 829 AAAAAACATTAGAAAAAGCGTTGCATGAGTTCGAAATATATGAACATTAAAAATCTG 888  
Qy 735 SerGlyLeuCyLysValAenThrLysAenAenSerGlyLysSerAapGluAenPalAry 754  
Db 889 CTTAGTTATTGTGAGAGAAATGACAAAGGAACT----- 921  
Qy 755 LysGluLeuCyGluLysLeuValGluGluValGluGluGlnCyLysValAleuProThr 774  
Db 922 -----GTTTGGCAAAATTTAGTGAATACTAAAAAGAGATGCTTCTTAAAGAAC 975  
Qy 775 GluLeuGlnGlnProAlaAlaAapLeuLysLysAapTyrLysThrTyrGluGluLys 794  
Db 976 GGAATGAATTAAGGAAAGAGAGACTTGACAAAGATGAAGAAAAAGATACGATCGCTTAA 1035  
Qy 795 LysAryAlaGluGluAlaMetAenLysSerSerLeuValLeuSerLeuLysLys--- 813  
Db 1036 AAGCGGCAGAAAGATCTACAAAGAAAGCTTATTTGCTATCAAGACCTAAGCAAACT 1095  
Qy 813 ----- 813  
Db 1096 GTAATGCCAAGTGGCGCAAGATGCAAGTCTTCAAGCAAGATATTACAAACGTAACA 1155  
Qy 814 -----AenGluSerAenValSerLysSerAenSerLysAenLysAapLysAenAlaVal 831  
Db 1156 GAATCAGGGTATCATCAGGGTCCGCATCATCACCAACAGGCGCACATCGACCAACA 1215  
Qy 832 SerAenGlyLeu-----GlnAapThrThr 839  
Db 1216 CAAAAATGGAACGCCAGCCACACAGTGAAGCACACAGGCAACCAAGCAAGTGAACGACG 1275  
Qy 840 LysHisValLys-----IleLeuAryAryGlyValLysAapValSerValThrGluLeu 857  
Db 1276 GGGCCCTGCAAAACTTGACTGTTAAAGACATATGTACTGAAGAGATATCAGAAAGCA 1335  
Qy 858 GluAlaLysAlaPheAapLeuAlaAlaGluValPheGlyAryValAapLeuLysGlu 877  
Db 1336 GAGGTCAAAGATTTGATGCAACAAACATGACATTTGAGATTGGAATTAAGAA 1395  
Qy 878 ArgCyAenLysLeuGluSerAapCyAryGlyLysGluAapCyLysAapLeuGluGlu 897  
Db 1396 GAATGTAAGCCTTATGAAATTTGCGGTTTAAAGAGATTTGAAGAAACTGAACCA 1455  
Qy 898 ValGlyLysLysIleAenLysAlaCyAryAenLysLysProLeuGluValLysProHis 917  
Db 1456 GCTTGTAAAGAAATGAAAGATTTATGTAA--CTGGAAGCATTAAGAGTTGCGCTCAT 1512  
Qy 918 -----GluThrValThrGluSerThrThrThrThrThrThrThrVal--- 934  
Db 1513 CATACAGAGACATTAACAAATTAAGTGAACGAAACACAGACGAAACAAAGACCGTTGAG 1572  
Qy 935 ---AlaAapProLysAla----- 939  
Db 1573 AAGTCGATGACAAAGGCTGATGTAAGACCGTTGAAGAACTGTTACGTAACCAACA 1632

Qy 940 -----ThrGluCyLysSerLeuGlnThrThrAapThrTyrVal 952  
Db 1633 GGAAGTGAAGAAAAAGTAAACAGAAAGTGTACATGATACAAACACAGATATCGGGTG 1692  
Qy 953 ThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIle 972  
Db 1693 ACAAGCAGCTTCATTTGCAATACGATACGACCAACAGATGATCGACGGTGAACAGTG 1752  
Qy 973 ThrLeuThrSerThrAryAryCyLysProThrLysCyThrThr-----GlyGluGlu 990  
Db 1753 ACGTTGACCTCGATGCGCAAGTGCAGCCTTACCAATGTACTGATTCAGACAGAGAG 1812  
Qy 991 AapAapAlaGly-----AapValLysProSerGluGlyLeuAryMetSer 1005  
Db 1813 ACAGATTAAGAGAGAGAGAGAGAGAAAGATGTAAACCGAATGACGAGATGAATAAG 1872  
Qy 1006 GlyTyrAenValMetAryGlyValIleValAlaMetValIleSerPheMetIle 1023  
Db 1873 GTTCTGATATGATTAATAATATGTTGTTGCGAGTGTATTGCGAATGATG 1926

Search completed: October 29, 2005, 10:12:11  
Job time : 1430 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_blue\_p2n model

Run on: October 29, 2005, 08:57:56 ; Search time 7490 Seconds  
(without alignments)  
5198.898 Million cell updates/sec

Title: US-10-654-416-14  
Perfect score: 5411  
Sequence: 1 MARAVKQAGTQNSIDEH.....MSGNNVGRGVIVAVISFMI 1023

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10654416/funac\_26102005\_101023\_21745/app\_query.fasta\_1.1159  
-DB=EST -QFMT=fastap -SUPFX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -XGAPOP=200000000  
-USER=US10654416.@CGN 1.1.5514.@funac\_26102005\_101023\_21745 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_g881: \*  
9: gb\_g882: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	460.5	8.5	746	2	AM334782 S37H4 AGS
C 2	455	8.4	789	2	AM335265 S45B7 AGS
C 3	453.5	8.4	703	2	AM331960 S1B8 AGS-
C 4	450.5	8.3	736	2	AM333066 S16G12 AG
C 5	443.5	8.2	693	2	AM334544 S36B1 AGS
C 6	441.5	8.2	742	2	AM335497 S47H6 AGS
C 7	430	7.9	737	2	AM334880 S4OC12 AG
C 8	429	7.9	725	2	AM333034 S16D10 AG
C 9	427	7.9	705	2	AM334533 S36A7 AGS

C 10	415	7.7	719	2	AM334763	AM334763 S37F11 AG
C 11	409.5	7.6	647	2	AM334174	AM334174 S31D8 AGS
C 12	402	7.4	649	2	AM331911	AM331911 S2C12 AGS
C 13	394	7.3	733	2	AM332556	AM332556 S9H7 AGS-
C 14	393	7.3	710	2	AM332531	AM332531 S9E12 AGS
C 15	393	7.3	800	2	AM334916	AM334916 S40G3 AGS
C 16	389	7.2	572	2	AM335173	AM335173 S44A3 AGS
C 17	387.5	7.2	699	2	AM333354	AM333354 S20F5 AGS
C 18	384	7.1	617	2	AM332662	AM332662 S11C9 AGS
C 19	380.5	7.0	593	2	AM333429	AM333429 S21P6 AGS
C 20	376.5	7.0	739	2	AM334513	AM334513 S35G10 AG
C 21	374.5	6.9	428	2	AM334345	AM334345 S33E10 AG
C 22	367	6.8	678	2	AM331939	AM331939 S2P3 AGS-
C 23	364	6.7	739	2	AM334475	AM334475 S35C11 AG
C 24	363	6.7	654	2	AM334208	AM334208 S31G2 AGS
C 25	363	6.7	761	2	AM334463	AM334463 S35B3 AGS
C 26	362.5	6.7	743	2	AM334196	AM334196 S31P5 AGS
C 27	359.5	6.6	694	2	AM334689	AM334689 S36G7 AGS
C 28	358	6.6	714	2	AM334897	AM334897 S40B6 AGS
C 29	350.5	6.5	566	2	AM332050	AM332050 S3E3 AGS-
C 30	343	6.3	716	2	AM334785	AM334785 S37H11 AG
C 31	331	6.1	540	2	AM334225	AM334225 S32A3 AGS
C 32	324.5	6.0	682	2	AM335277	AM335277 S45C1 AGS
C 33	318	5.9	559	2	AM333479	AM333479 S22C3 AGS
C 34	317	5.9	576	2	AM333182	AM333182 S18D8 AGS
C 35	316	5.8	582	2	AM332456	AM332456 S8F12 AGS
C 36	312	5.8	624	2	AM334653	AM334653 S38G9 AGS
C 37	307	5.7	658	2	AM334501	AM334501 S35P7 AGS
C 38	304	5.6	684	2	AM333103	AM333103 S17C12 AG
C 39	304	5.6	572	2	AM3332158	AM3332158 S4H11 AGS
C 40	300.5	5.6	527	2	AM334227	AM334227 S32A8 AGS
C 41	298.5	5.5	563	2	AM334815	AM334815 S39E9 AGS
C 42	297.5	5.5	743	2	AM335500	AM335500 S47H10 AG
C 43	297	5.5	603	2	AM332931	AM332931 S15A5 AGS
C 44	292	5.4	575	2	AM332985	AM332985 S15G12 AG
C 45	286	5.3	612	2	AM334018	AM334018 S29C10 AG

#### ALIGNMENTS

RESULT 1  
AM334782/c 746 bp mRNA linear EST 31-JAN-2000  
LOCUS S37H4 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.  
DEFINITION AM334782  
ACCESSION AM334782  
VERSION AM334782.1 GI:6831048  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Pneumocystis carinii  
Pneumocystis carinii  
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE  
1 (bases 1 to 746)  
Smullan,A.G., Arnold,V., Weise,M., Wunderlich,J., Staben,C.,  
Edman,J.C., Kovacs,J. and Cuelhion,M.  
Expressed sequence tags from Pneumocystis carinii  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu.  
Location/Qualifiers  
1..746  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_id="AGS-1"  
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;  
P. carinii organisms (3x10e9) from a single rat (99-1-6,

FEATURES  
source





```

Db      251  ACCACCGATTCAAAACAAGACAGATTAAGAGAGAGAGAGAGAGAGATGTAA 192
Qy      998  ProSerGlyLeuArgMetSerGlyTTPAsnValMetArgGlyValIleValAlaMet 1017
Db      191  CCAAAATGATGGGATGAAATAAGAGTTCTGATATGATTAATAATATGTTGGAGTG 132
Qy      1018  ValIleSerPheMetIle 1023
Db      131  ATTGTATGGGAGATGATG 114

RESULT 3
LOCUS   AM331960 703 bp mRNA linear EST 31-JAN-2000
DEFINITION SIB8 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM331960
VERSION   AM331960.1 GI:6828226
KEYWORDS  EST.
SOURCE    Pneumocystis carinii
           Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
           Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 703)
           Smullan,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
           Edman,J.C., Kovacs,J. and Cushion,M.
           Expressed sequence tags from Pneumocystis carinii
           Unpublished (2000)
JOURNAL  Contact: Staben C
           School of Biological Sciences
           University of Kentucky
           101 Morgan Building, University of Kentucky, Lexington, KY
           40506-0225 USA
           Tel: 606 257 2161
           Fax: 606 257 1717
           Email: staben@pop.uky.edu.
           Location/Qualifiers
FEATURES
     source          1..703
                     /organism="Pneumocystis carinii"
                     /mol_type="mRNA"
                     /db_xref="taxon:4754"
                     /lab_host="E. coli"
                     /clone_lib="AGS-1"
                     /note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
                     P. carinii organisms (3x10e9) from a single rat (99-1-6,
                     sacrificed on 3/17/99) at Cincinnati VA facilities.
                     Trizol extracted RNA. Oligo dt priming, standard
                     conditions described by vendor, Stratagene. Further
                     details see www.uky.edu/Project/Pneumocystis/"
ORIGIN
Alignment Scores:
Pred. No.:      3.95e-35      Length:      703
Score:          453.50      Matches:      92
Percent Similarity: 63.00%      Conservative: 34
Best Local Similarity: 46.00%      Mismatches: 41
Query Match:    8.38%      Indels:      33
DB:             2           Gaps:      5

US-10-654-416-14 (1-1023) x AM331960 (1-703)
Qy      854  ValThrGluLeuGluValAlaValPheAspLeuAlaIleGluValPheGlyArgTyrVal 873
Db      688  GTATCAAGAGCAAGAAAGTAAAGCATTTGATGCAACGACGATGACATTCATTTTG 629
Qy      874  AspleuLygGluArgCyAsnLyLeuGluSerAspCyAsArgIleLygGluAspCyLyAs 893
Db      628  GAATTAAGAGCAATCAATGCTTTAGAACTAGATTGCGGTTTAAAGAGATTGTCA 569
Qy      894  AspleuGluGluValCyGlyLyLyIleAsnLyValAcGAsArgAsnLeuLyProLeuGlu 913
Db      568  GATCTT---CAAGTTTCGGAGAAATAGACAAAGTTATGCAAAAGAAACCATTTGAA 512
Qy      914  ValLyProHisGluThrValThrGlu-----SerThrThrThrThrThrThrThr 930

```

```

Db      511  ATTAAGCTCATCATCATACAGAGACGAAAAAGAAATCTCAACCACTACGACGACCACTACG 452
Qy      931  ThrThrThrValAlaAspProLySAlaThr----- 940
Db      451  ACCAGCACTACCAACAACACTACTACGACTACTACGACTACTACTACTACTACTACTACTACT 352
Qy      941  -----GluCyLySerLeuGlnThrThrAspThrTTPVal 952
Db      391  CGAAGTCAGAGAAACATAACAGAGAGTGTACATATACATACAGACAGACAGATGATGCGT 332
Qy      953  ThrGlnThrSerThrIleThrSerThrIleThrSerThrIleThrSerLyIle 972
Db      331  ACAAGTACGTCATTCATGATACGATACGACAAACAGTACGTCGACAGTACGTCAGACAGTG 272
Qy      973  ThrLeuThrSerThrArgArgCyLySerProThrLyCyThrThr----- 987
Db      271  ACGTTACGTCGATGTCAGTGCACAACTACCAAAATGTACCAACGATTCACAAACAAAGAG 212
Qy      988  -----GlyGluGluAspAspAlaGlyAspValLyProSerGlyLeuArg 1003
Db      211  ACAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158
Qy      1004  MetSerGlyTTPAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db      157  ATTAAGATTCCTGATATGATTAATAATATGTTGAGATGATTTATGCGGAGATGATG 98

RESULT 4
LOCUS   AM333066 736 bp mRNA linear EST 31-JAN-2000
DEFINITION S16G12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM333066
VERSION   AM333066.1 GI:6829423
KEYWORDS  EST.
SOURCE    Pneumocystis carinii
           Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
           Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 736)
           Smullan,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
           Edman,J.C., Kovacs,J. and Cushion,M.
           Expressed sequence tags from Pneumocystis carinii
           Unpublished (2000)
JOURNAL  Contact: Staben C
           School of Biological Sciences
           University of Kentucky
           101 Morgan Building, University of Kentucky, Lexington, KY
           40506-0225, USA
           Tel: 606 257 2161
           Fax: 606 257 1717
           Email: staben@pop.uky.edu.
           Location/Qualifiers
FEATURES
     source          1..736
                     /organism="Pneumocystis carinii"
                     /mol_type="mRNA"
                     /db_xref="taxon:4754"
                     /lab_host="E. coli"
                     /clone_lib="AGS-1"
                     /note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
                     P. carinii organisms (3x10e9) from a single rat (99-1-6,
                     sacrificed on 3/17/99) at Cincinnati VA facilities.
                     Trizol extracted RNA. Oligo dt priming, standard
                     conditions described by vendor, Stratagene. Further
                     details see www.uky.edu/Project/Pneumocystis/"
ORIGIN
Alignment Scores:
Pred. No.:      8.57e-35      Length:      736
Score:          450.50      Matches:      88
Percent Similarity: 63.16%      Conservative: 44
Best Local Similarity: 42.11%      Mismatches: 48
Query Match:    8.33%      Indels:      29
DB:             2           Gaps:      4

```

US-10-654-416-14 (1-1023) x AM333066 (1-736)  
 QY 844 ILELEUARGAGGlyValLysAspValSerValThrGluLeuGluAlaLysAlaIleAsp 863  
 Db 725 CTCGTTAAAGAGCGATTATGTATGAGAGAGATGATGAAGCAGAGATGAAGCATTTAT 666  
 QY 864 LeuAlaIaIaGluValPheGlyYArgTyValAspLeuLysGluArgCysAsnLysLeuGlu 883  
 Db 665 GCAGCGCCCGGTAGCATTTGGAGATGATATTTTGGAATTTGAAGAAGAGATGACGCCCTTTACA 606  
 QY 884 SerAspCysArgLLeuGluLysCysLysAspLeuGluGluValCysLysLysIleAsn 903  
 Db 605 CTGATGTCGGGTTTGAAGAGAGAAATGTCAGGAATCAACAAGCTTCGGAAAAATATGAC 546  
 QY 904 LysAlaCysArgAsnLeuLysProLeuGluValLysProHISgluThrValThrIcu--- 922  
 Db 545 GAGTTATCGGAGGAATATAAACCATTAAGAAATTAAAGCCTCATCATACAGAGCGGAAA 486  
 QY 923 -----SerThrThrThrThr 927  
 Db 485 GAAATCTCAACCACTAGCAGACGACCACTAGCAGACTACCAACAACACTACCAACT 426  
 QY 928 ThrThrThrThrThrValAlaAspPro-----LysAlaThr---GluCys 942  
 Db 425 ACGACGAACTACTACTACTCAACAACCAAGCCGGAGAGTGAGGAAAAAGTAAACGAAGAGTGT 366  
 QY 943 LysSerLeuGlnThrThrAspThrThrValThrGlnThrSerThrHisThrSerThrSer 962  
 Db 365 ACAATGATACAAACAACAGATACATGAGGTAAAGCGGTACGTCTTCATACAGAGTACGACA 306  
 QY 963 ThrIleThrSerThrIleThrSerLysIleThrIleuThrSerThrArgArgCysLysPro 982  
 Db 305 ACGAGTACGTGCGACAGTGCACGTGACAGTGAAGTTCGATCGATGCGAAGTCAAGCCT 246  
 QY 983 ThrLysCysThrThr-----GlyGluGluAspAlaIcu 994  
 Db 245 ACCAAATGTACCAACGATTCAAACAAGACAGATAAAGAGGAGAAAGAAAGAGAGAA 186  
 QY 995 AspValLysProSerGluGlyLeuArgMetSerGlyTPAsnValMetArgGlyValIle 101  
 Db 185 GATGTAAACCGAATGATGGATGGATGAAATTAAGAGTTCCTGATATGATTAATATGTTG 126  
 QY 1015 ValAlaMetValIleSerPheMetIle 1023  
 Db 125 TTGGGAGTGAATTGTATGGGAGTGATG 99  
 RESULT 5  
 AM334544/c  
 LOCUS 693 bp mRNA linear EST 31-JAN-2000  
 DEFINITION SJB1 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.  
 ACCESSION AM334544  
 VERSION AM334544.1 GI:6830901  
 KEYWORDS EST.  
 SOURCE Pneumocystis carinii  
 ORGANISM Pneumocystis carinii  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.  
 REFERENCE 1 (bases 1 to 693)  
 AUTHORS Smilian,A.G., Arnold,J., Wise,M., Wunderlich,J., Staben,C.,  
 Edman,J.C., Kovacs,J. and Cushion,M.  
 TITLE Expressed sequence tags from Pneumocystis carinii  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Staben C  
 School of Biological Sciences  
 University of Kentucky  
 101 Morgan Building, University of Kentucky, Lexington, KY  
 40506-0223, USA  
 Tel: 606 257 2161  
 Fax: 606 257 1717  
 Email: staben@pop.uky.edu.  
 FEATURES  
 CDS  
 location/Qualifiers  
 1..693

```

/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_id="AGS-1"
/notes="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Tiziol extracted RNA. Oligo dT priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/project/pneumocystis/"

ORIGIN

Alignment Scores:
Pred. No.:          4.01e-34          Length:          693
Score:              443.50           Matches:          91
Percent Similarity: 61.24%           Conservative:     37
Best Local Similarity: 43.54%         Mismatches:      48
Query Match:        8.20%            Indels:          33
Db:                  2                Gaps:            5

US-10-654-416-14 (1-1023) x AM3354544 (1-693)

Oy      845  leuArgArgGlyValIysAspValSerValThrGluLeuGluAlaIysAlaPheAspLeu 864
      :::::|||||
Db      692  GTTAAAGGGGATATGTAGCTGAGGAGGTATCTCAAGCAGAAAGTAAAGCATTTGATGCA 633
      |||
Oy      865  AlAlaGluValPheGlyArgTyrValAspLeuIysGluArgCysAsnLysLeuGluSer 884
      |||
Db      632  ACGACGATGACATGTGAGCTGTATTTGGATTGAAAGAGAAATGCATTCCTTAGAACTA 573
      |||
Oy      885  AspCysArgIleLysGluAspCysLysAspLeuGluGluValCysLysIleAsnLys 904
      |||
Db      572  GATTGCGGCTTTTAAAGGAGATTGTCGAGCTT---CAAGTTTGGCGGAGAAATGTCAG 516
      |||
Oy      905  AlaCysArgAsnLeuLysProLeuGluValIysProHisGluThrValIThrGlu----- 922
      |||
Db      515  TTAATCAAGAAATAGAACCATTTAGAAATTTAAGCTCATCATACAGAGAGCAAAAAAGAA 456
      |||
Oy      923  -----SerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 930
      |||
Db      455  ATCTCAACCACTACGCGGACCACTACGACGACCACTACCAACAACCTACTACGACACTACT 396
      |||
Oy      931  ThrThrThrValAlaAspProLys-----AlaThrGluCysLys 943
      |||
Db      395  ACGACTACTACTACGACCAACCAAGCAGAGAAATGGAGAAAAGTAAACAGAAAGACTGTACA 336
      |||
Oy      944  SerLeuGlnThrThrAspThrTrpValIThrGlnThrSerThrHisThrSerThrSerThr 963
      |||
Db      335  ATGATACAGACGACGATACATGCGGTACAAAGTACGTCATTCGATACGATGACGACAAACG 276
      |||
Oy      964  IleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArgCysLysProThr 983
      |||
Db      275  AGTAGCGTCGACAGGACGTGACAGTACGTCGATCGTCGATGCGCAAGTCGACAAACCTAAC 216
      |||
Oy      984  LysCysThrThr-----GlyGluGluAspPalaGly 994
      |||
Db      215  AAAAGTACCAACCGATTCMAACAAAGACAGATTAAGAGAGAGAGAAAGAAAGAA----- 162
      |||
Oy      995  AspValIysProSerGluGluLysArgMetSerGlyTrpAsnValMetArgGlyValIle 1014
      |||
Db      161  GAACTAAACCAAAATGATGGATATAAATATAGACTTCCTGATATGATTAATAATATGTG 102
      |||
Oy      1015  ValAlaMetValIleSerPheMetIle 1023
      |||
Db      101  TTGGGAGTGATTTTATGGGAGTATG 75
      |||

RESULT 6
LOCUS      AM335497/c
DEFINITION  742 bp  mRNA
ACCESSION   AM335497
VERSION     AM335497.1  GI:6832118

```

KEYWORDS EST.  
SOURCE Pneumocystis carinii  
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 742)  
Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,  
Edman,J.C., Kovacs,J. and Cushman,M.  
TITLE Expressed sequence tags from Pneumocystis carinii  
JOURNAL Unpublished (2000)  
COMMENT Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu.  
Location/Qualifiers  
1. 742  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_lhb="AGS-1"  
/note="Vector: Lambda ZAP II; Site\_1: EcoRI; Site\_2: XhoI;  
P. carinii organisms (3x10e9) from a single rat (99-1-6,  
sacrificed on 3/17/99) at Cincinnati VA facilities.  
Trizol extracted RNA. Oligo dt priming, standard  
conditions described by vendor, Stratagene. Further  
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:  
Pred. No.: 7,13e-34 Length: 742  
Score: 441.50 Matches: 92  
Percent Similarity: 60.68% Conservative: 33  
Best Local Similarity: 44.66% Mismatches: 50  
Query Match: 8.16% Indels: 31  
Gaps: 5

US-10-654-416-14 (1-1023) x AW335497 (1-742)

QY 847 ArgGlyValIysAspValSerValThrGluLeuGluAlaValAspAlaPheAspLeuAla 866  
DB 741 AGACCATATGATGATGAGAGATCAGAACCAAGCATTTGACGCAACAGC 682  
QY 867 GluValPheGlyArgValAspLeuLysGluArgCysAsnLysLeuGluSerAspCys 886  
DB 681 GTACCATTTGCAATTTGTAATTTGAAAGAGAAATGTAAGCTTTAGAAATTGATTGC 622  
QY 887 ArgGlyLeuGluAspCysValAspLeuGluValCysLysLysLeuAsnLysAlaCys 906  
DB 621 GGTTTTAAAGAGATTTGTCAGATCTT---CAAGTTTGTGAGAAATAGACGTTATGC 565  
QY 907 ArgAsnLeuLysPheLeuGluValLysPheHisGluThrValThrGlu-----Ser 923  
DB 564 GGAGCAATTAACCAATTAAGAAATTAAGCTCATATACAGACGCAAAAAGAAATCTCA 505  
QY 924 ThrThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThr----- 940  
DB 504 ACCCACTACGACGACCACTACGACGACCTACCACTACTACGACGACGACGACA 445  
QY 941 -----GluCysLysSerLeu 945  
DB 444 ACAACTACTACTACCAAGCAGGAAAGTGGAGAAAGTAAGACAGAAAGATGATACATGATA 385  
QY 946 GlnThrThrAspThrTrpValThrGlnThrSerThrHisThrSerThrSerThrIleThr 965  
DB 384 CAAACCAACAGATACATGGTAAACGATGATTCATTCATGATGATGATGATGATGATGATG 325  
QY 966 SerThrIleThrSerValIleThrLeuThrSerThrArgAspCysLysPheThrLysCys 985

DB 324 TCAACGTCGACGTGACAGTGAAGCTTGACGTGCATGAGAAAGTGCACCAACTACCAATGT 265  
QY 986 ThrThr-----GlyGluGluAspAlaGlyAspValLys 997  
DB 264 ACCACCGATTCACAGCAGACAGATTAAGGAGAGAAATGAACA---GAAGTAA 208  
QY 998 ProSerGluLysLeuArgMetSerGlyTrpAsnValMetArgGlyValIleValMet 1017  
DB 207 CCAATGATGATGATGAATAAATTAAGAGTTCCGTATATATTAATAATATGTTGGAGTGT 148  
QY 1018 ValIleSerPheMetIle 1023  
DB 147 ATTGTTATGCAATGATGAT 130

RESULT 7  
AW334880/c 737 bp mRNA linear EST 31-JAN-2000  
LOCUS AW334880  
DEFINITION s40c12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.  
ACCESSION AW334880  
VERSION AW334880.1 GI:6831237  
KEYWORDS EST.  
SOURCE Pneumocystis carinii  
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.  
REFERENCE 1 (bases 1 to 737)  
Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,  
Edman,J.C., Kovacs,J. and Cushman,M.  
TITLE Expressed sequence tags from Pneumocystis carinii  
JOURNAL Unpublished (2000)  
COMMENT Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu.  
Location/Qualifiers  
1. 737  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_lhb="AGS-1"  
/note="Vector: Lambda ZAP II; Site\_1: EcoRI; Site\_2: XhoI;  
P. carinii organisms (3x10e9) from a single rat (99-1-6,  
sacrificed on 3/17/99) at Cincinnati VA facilities.  
Trizol extracted RNA. Oligo dt priming, standard  
conditions described by vendor, Stratagene. Further  
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:  
Pred. No.: 1.04e-32 Length: 737  
Score: 430.00 Matches: 93  
Percent Similarity: 56.35% Conservative: 49  
Best Local Similarity: 36.90% Mismatches: 96  
Query Match: 7.95% Indels: 14  
Gaps: 5

US-10-654-416-14 (1-1023) x AW334880 (1-737)

QY 553 CysHisThrLeuGluGluGlnCysAsnArgLeuGlyThrThrGluIleLeuLysGlnVal 572  
DB 737 TGATTAATTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 678  
QY 573 LeuLeuAspGluHisLysAspThrThrLeuLysAspGlnGlnSerCysValLysTrpLeuLys 592  
DB 677 TTTTAAAGAGAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 618  
QY 593 GluLysCysAsnLysTrpSerArgArgGlyAspAspArgPheSerPheValPhe 612



JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: staben@pop.uky.edu

FEATURES  
Location/Qualifiers

1..705  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_lib="AGS-1"  
/note="Vector: Lambda ZAP II; Site\_1: EcoRI; Site\_2: XhoI;  
P. carinii organisms (3x10e9) from a single rat (99-1-6,  
sacrificed on 3/17/99) at Cincinnati VA facilities.  
Rizol extracted RNA. Oligo dt priming, standard  
conditions described by vendor, Stratagene. Further  
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:

Pred. No.: 1.95e-32 Length: 705  
Score: 427.00 Matches: 86  
Percent Similarity: 61.03% Conservative: 33  
Best Local Similarity: 44.10% Mismatches: 46  
Query Match: 7.89% Indels: 30  
DB: 2 Gaps: 4

US-10-654-416-14 (1-1023) x AM334533 (1-705)

858 GluAlaValAlaPheAspLeuAlaGluValPheGlyArgTyrValAspLeuLysGlu 877  
702 GAGGTAAAGCATTTGATGCAACGACGATGATTTGCAATTGGAATTGAAAGAA 643  
878 ArgCysAlaLeuLysGluSerAspCysArgTyrLeuGluAspCysLeuGlu 897  
642 GAATGTAAAGCTTTAGAACTAGATTCGCTTTAAAGATTTGCGAGTGTGATGAT 583  
898 ValCysLysLysIleAsnLysAlaCysArgAsnLeuLysPheLeuGluValLysPhe 917  
582 GTTTCGAAAGAAATTGACACTTATGTAA--CTGAAACATTGAAATTAACCTCAT 526  
918 GluThrValThrGlu-----SerThrThrThrThrThrThrThrThrVal 934  
525 CATACAGAGAGCAAAAGAAATCTCAACCACTACGACGACCACTACGACGACTAAC 466  
935 AlaAspProLysAlaThr----- 940  
465 ACAACTACTACGACGACGACGACCAACTACTACTACGACGACGAGAGTGAAGA 406  
941 -----GluCysLysSerLeuGlnThrThrAspThrTrpValThrGlnThrSer 956  
405 AAAGTAACAGAGAGTGTACATATACACGACGACGACGATGATGAGTGAAGCACTCA 346  
957 ThrIleSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSer 976  
345 TTGCATACGAGTACGACAGAGTACGTCAGACGACGTCGACAGTGAAGTTCACCTCG 286  
977 ThrIleArgCysLysPheProThrLysCysThr-----ThrGly 988  
285 ATGAGAAAGTGCACCAATGTACCCCGATTCCAAACAAAGACAGATTAAGGA 226  
989 GluGluAspAlaGlyAspValLysPheSerGluGlyLeuArgMetSerGlyTrpAsn 1008  
225 GAAAGAGAGAGAGAAAGATGTAAACCAATGATGAGATGAATTAAGAGTTCCGAT 166  
1009 ValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023  
165 ATGATTAATAATGATGTTGCGAGTGTATGCGGATGATG 121

RESULT 10  
AM334763/c

LOCUS AM334763 719 bp mRNA linear EST 31-JAN-2000

DEFINITION S37F11 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.

ACCESSION AM334763

VERSION AM334763.1 GI:6831029

KEYWORDS EST.

SOURCE Pneumocystis carinii

ORGANISM Pneumocystis carinii

REFERENCE Bukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
Pneumocystidaceae; Pneumocystis.

AUTHORS 1 (bases 1 to 719)  
Smullen, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,  
Edman, J.C., Kovacs, J. and Cushion, M.

Expressed sequence tags from Pneumocystis carinii

Unpublished (2000)

Contact: Staben C

School of Biological Sciences

University of Kentucky

101 Morgan Building, University of Kentucky, Lexington, KY

40506-0225, USA

Tel: 606 257 2161

Fax: 606 257 1717

Email: staben@pop.uky.edu.

Location/Qualifiers

1..719  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_lib="AGS-1"  
/note="Vector: Lambda ZAP II; Site\_1: EcoRI; Site\_2: XhoI;  
P. carinii organisms (3x10e9) from a single rat (99-1-6,  
sacrificed on 3/17/99) at Cincinnati VA facilities.  
Rizol extracted RNA. Oligo dt priming, standard  
conditions described by vendor, Stratagene. Further  
details see www.uky.edu/Project/pneumocystis/"

ORIGIN

Alignment Scores:

Pred. No.: 3.34e-31 Length: 719  
Score: 415.00 Matches: 95  
Percent Similarity: 53.76% Conservative: 48  
Best Local Similarity: 35.71% Mismatches: 88  
Query Match: 7.67% Indels: 35  
DB: 2 Gaps: 6

US-10-654-416-14 (1-1023) x AM334763 (1-719)

60 LysLeuLysGlyPheCysGluAsp---GlyLysAlaAspThrLysCysLysGluLeuLys 78  
712 AGAATTAAAGATTGTGTAAGATACGAAAACTTAAGAAAATGTCTGCTATTAAG 653  
79 AlaAsnIleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLys 98  
652 ACAGAGTTTAAGAAAAGTTGAGACGAAATGAAAGTCCATTATGATCATCTTTCAAT 593  
99 LysIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeu 118  
592 CCT-----TCAGATGAGAAATGCGCAGAAATATGACCAAAATGTTTATTTTA 545  
119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuAsnLysCysTyr 138  
544 AGC----- 542  
139 GlnLys-LysArgAspLysValAlaGluGluValIleLeuLeuArgAlaLeuArgSerAspLe 158  
541 CGAAGCAACGCGTCAAAATGTGCGAGATGAAATTTCTTTTAAAGCTTTTAAAGAACAGT 482  
158 uAsnGlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgG1 178  
481 TAAATGACACGATTAATGATTAAGAAAGATGAAAGAGTTTGTCCCAATATTAAATAGAGC 422

```

Oy uSeAspdluLeuThraenLeuCySleuAsglnlysgluThrCySlyAsnIleLeuI 198
Db 178 421 GAGGACAGCTTAATGTTTCATGATATTATTTTAAACTACAGTGTGAATCTCGAAA 362
Oy 198 eGluLyAspLySlyCySgLyThrLeuLyThraSpValSerIaIaLeuGlySerPh 218
Db 361 AAAAAAGATATTTTTCGAAACCTTTTAAAGCAAAATTTGAGTGGAG-----AAAGA 311
Oy 218 eLyLyLsgIuThrCySleuGluLeuLeuGluGlnCyStryPhetyrIleGlyAsnCySgl 238
Db 310 ATTGAGGGAAGATGTCATGAAAGGTTGGAAAAAGTTATTTTAAATCAAGTGGCC 251
Oy 238 yASpAspApleIleLySyrIleGluLeuGlyLyCySgIngluGlnAsnIleAl 258
Db 250 TAAATCAGAG-----TGTGATTAATATTAAAAAGCAATCGAGAAAAAATATTAC 200
Oy 258 ATyTMeTrProGclYProAspPhAspProThrxTrgProGluIaThrIleAlaGluAs 278
Db 199 ATACAAGCTTCAGAGATCAGATTTTCTCCAGTTCAACCAAAACCAATATGTTAGAGA 140
Oy 278 pIleGlyLeuGluGlnuPhetyrLySlyValGluGluAspGlyValIpheIleGlyAsAs 298
Db 139 GATTGGCTGGAAAGAAATTATTAAAGAAGAGAAATAACGAATTATGTCGGAAATCC 80
Oy 298 nHleuLyrgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLyLy 318
Db 79 AGGAGAAAAGAAAGTTGGAGATATATCTATTATTATG-----AGTAAAAAAA 32
Oy 318 eLyAspAspLySlyGlu 323
Db 31 AAAAAAAAAACTCGAG 16

```

RESULT 11  
 LOCUS AM334174/c  
 DEFINITION S31D8 AGS-1 *Pneumocystis carinii* CDNA 3', mRNA sequence.  
 ACCESSION AM334174  
 VERSION AM334174.1  
 KEYWORDS GI:6630531  
 EST.  
 SOURCE *Pneumocystis carinii*  
 ORGANISM *Pneumocystis carinii*  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.  
 REFERENCE 1 (bases 1 to 647)  
 Smallin,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,  
 Edman,J.C., Kovacs,J. and Cushion,M.  
 TITLE Expressed sequence tags from *Pneumocystis carinii*  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Staben C  
 School of Biological Sciences  
 University of Kentucky  
 101 Morgan Building, University of Kentucky, Lexington, KY  
 40506-0225, USA  
 Tel: 606 257 2161  
 Fax: 606 257 1717  
 Email: staben@pop.uky.edu  
 FEATURES  
 source location/Qualifiers  
 1..647

ORIGIN

Alignment Scores:

Pred. No.:	1,02e-30	Length:	647
Score:	409.50	Matches:	82
Percent Similarity:	63.1%	Conservative:	33
Best Local Similarity:	45.05%	Mismatches:	34
Query Match:	7.57%	Indels:	33
DB:	2	Gaps:	5

US-10-654-416-14 (1-1023) X AW334174 (1-6477)

Oy	872	TyValAspLeuLysGluAspGlyAsnLysLeuGluSerAspCysArgGlyLeuGluAsp	891
Db	639	TATTGGAAATTGAAAGAGCAATGCATGCTTTAGAACTAGAGTCCGGTTTAAAGAGAT	580
Oy	892	CysLysAspLeuGluGluValCysLysLysLysLeuLysAlaCysArgAsnLeuLysPro	911
Db	579	TGTTCAAGATCTT---CAAGTTTGGCGAAGAAATAGCAAGTATGCAAAAGAAATAAACCA	523
Oy	912	LeuGluValLysProHisGlu-----	918
Db	522	TTAAATAATTAAAGCTCGTCATGCAGAGACGCAAAAGAAATCTCAACCTACGACGACC	463
Oy	919	---ThyValThrGluSerThrThrThrThrThrThrThrThrThrValAlaAspPro	937
Db	462	ACTACGGCGACCACTACCACTACTACTACGACTACTACGACTACTACTACGACAACC	403
Oy	938	Lys-----AlaThrGluCysLysSerLeuGluThrThrAspThr	950
Db	402	AAGCAGGGAAGTGGAGGAAAGTAACAGAGAGTGTAACAATGATATACAGACAGACAGATACA	343
Oy	951	TrpValThrGluThrSerThrHisThrSerThrIleThrSerThrIleThrSer	970
Db	342	TGGGTGCAAGTACGTATTGCATACACAGTACGCAACGAGTACGTCCACAGTGCACGTG	283
Oy	971	LysIleThrLeuThrSerThrArgAspCysLysProThrLysCysThrThr-----	987
Db	282	ACAGTGCAGTTGACGTGCATGCGGCAAGTGCACAACTTCCCAATGATCCACCGCATTTCAAC	223
Oy	988	-----GlyGluGluAspAspAlaGlyAspValLysProSerGluGly	1001
Db	222	AAAGAGACAGATTAAAGAGGAGGAAGAAAGAA-----GAAGTAAACCAAAATGATGGG	169
Oy	1002	LeuArgMetSerGlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPhe	1021
Db	168	ATGAAAATTAAGAGTCTCTGATATGATTAAATAATATGTTGTGGAGGAGTCATTGTATGGGG	109
Oy	1022	MetIle 1023	
Db	108	ATGATG 103	

RESULT 12	AM331911/c	649 bp	mRNA	linear	EST 31-JAN-2000
LOCUS	AM331911				
DEFINITION	S2312 AGS-1		Pneumocystis carinii	CDNA 3',	mRNA sequence.
ACCESSION	AM331911				
VERSION	AM331911.1	GI:6828320			
KEYWORDS	EST.				
SOURCE	Pneumocystis carinii				
ORGANISM	Pneumocystis carinii				
REFERENCE	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;				
AUTHORS	Pneumocystidaceae; Pneumocystis.				
	1 (bases 1 to 649)				
	Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,				
	Edman,J.C., Kovacs,J. and Cushion,M.				
TITLE	Expressed sequence tags from Pneumocystis carinii				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Staben C				
	School of Biological Sciences				
	University of Kentucky				
	101 Morgan Building, University of Kentucky, Lexington, KY				
	40506-0225, USA				
	Tel: 606 257 2161				
	Fax: 606 257 1717				
	Email: staben@pop.uky.edu.				



## FEATURES

Location/Qualifiers

1. 649  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_id="AGS-1"  
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI; P. carinii organisms (3x10e9) from a single rat (99-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming. Standard conditions described by vendor, Stratagene. Further details see www.uky.edu/project/Pneumocystis/"

## ORIGIN

## Alignment Scores:

Pred. No.: 5.95e-30 Length: 649  
Score: 402.00 Matches: 80  
Percent Similarity: 62.92% Conservative: 32  
Best Local Similarity: 44.94% Mismatches: 36  
Query Match: 7.43% Indels: 30  
DB: Gaps: 4

US-10-654-416-14 (1-1023) x AW331911 (1-649)

QY 875 LeuLysGluArgCysAenLysLeuGluSerAspCysArgIleLysGluAspCysLysAsp 894

DB 648 TTGAAAGAGAAAGTAAAGCTTTGAACTAGATGGCGTTTACAAAGAGATGTTCCAGT 589

QY 895 LeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluVal 914

DB 588 GTTGATGATGTTGCCAAGAAATTGACACATATGTTAA---CTGGAACCTTTGAATTT 532

QY 915 LysProHisGluThrValThrGlu-----SerThrThrThrThrThrThrThrThr 931

DB 531 AAGCTCATCATACAGAGACGCAAAAGAAATCTCAACCTACAGACGACATACAGACG 472

QY 932 ThrThrValAlaAspProLysAlaThr----- 940

DB 471 ACTACTACCACTACTACTACCAAGCTACGACGACCAACTACTACTACCAAGCAGGA 412

QY 941 -----GluCysLysSerLeuGlnThrThrAspThrTyrValThr 953

DB 411 AGTCGAGAGAAAGTAAAGAGAGATGTCATGATGACAGAGAGAGATGAGTGC 352

QY 954 GlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThr 973

DB 351 AGCAGCTCATTCGACGAGTACGACAAAGAGTACGTCGACGACGTCGACGAGTGC 292

QY 974 LeuThrSerThrArgArgCysLysProThrLysCysThrThr----- 987

DB 291 TTGACGTCGATGAGAAAGTGCAGCTTACCAAAAGTACACCGATTCAACAAAGAGACA 232

QY 988 -----GlyGluGluAspAspAlaGlyAspValLysProSerGluGlyLeuArgMetSer 1005

DB 231 GATAAAGAGAGAGAGAGAGAGAGAGATGTAAACCAATGATGGATGAATAAGA 172

QY 1006 GlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023

DB 171 GTTCTGATATGATTAAATATATGTTGCGAGATGTTATGGGAGTATG 118

RESULT 13

AW332556/c

LOCUS AW332556 733 bp mRNA linear EST 31-JAN-2000

DEFINITION S9H7 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.

ACCESSION AW332556

VERSION AW332556.1 GI:6828913

KEYWORDS EST.

SOURCE Pneumocystis carinii

ORGANISM Pneumocystis carinii

Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;

Pneumocystidaceae; Pneumocystis.

REFERENCE 1 (bases 1 to 733)

## AUTHORS

Smilian, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,  
Edman, J.C., Kovacs, J. and Cushion, M.  
Expressed sequence tags from Pneumocystis carinii  
Unpublished (2000)

## TITLE

Unpublished (2000)

## JOURNAL

Unpublished (2000)

## COMMENT

Unpublished (2000)

## FEATURES

Location/Qualifiers  
1. 733  
/organism="Pneumocystis carinii"  
/mol\_type="mRNA"  
/db\_xref="taxon:4754"  
/lab\_host="E. coli"  
/clone\_id="AGS-1"  
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI; P. carinii organisms (3x10e9) from a single rat (99-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming. Standard conditions described by vendor, Stratagene. Further details see www.uky.edu/project/pneumocystis/"

## ORIGIN

## Alignment Scores:

Pred. No.: 4.68e-29 Length: 733  
Score: 394.00 Matches: 84  
Percent Similarity: 60.20% Conservative: 37  
Best Local Similarity: 41.79% Mismatches: 46  
Query Match: 7.28% Indels: 34  
DB: Gaps: 6

US-10-654-416-14 (1-1023) x AW332556 (1-733)

QY 854 ValThrGluLeuGluAlaLysAlaPheAspLeuAlaGluValPheGlyArgTyrVal 873

DB 707 GTATCAAGACAGAGAGAGAGAGATGTCATGATGACACCAATAGCAAGTGTATTTG 648

QY 874 AspLeuLysGluArgCysAenLysLeuGluSerAspCysArgIleLysGluAspCysLys 893

DB 647 GAATTTGAAGAGAAATCAATGCTTTGAACTAGATGGCGTTTAAAGAGATGTCAG 588

QY 894 AspLeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGlu 913

DB 587 GAATCTAAACGCTTGTAAAGAAATGACACGTTATGCAAAAGAAATPAAACATTGAA 528

QY 914 ValLysProHis-----GluThrValThrGluSerThrThrThrThr----- 928

DB 527 GTTACGCTTATCTACAGAGACAAATATATAAAGAAACCATGACATGATGATGACGGA 468

QY 929 -----ThrThrThrThrValAlaAsp 936

DB 467 ACAAAAGCAATTGACAAAGCTGATGGAAGACCGGTACGAAGACTATATGACGAG 408

QY 937 Pro-----LysAlaThrGlu---CysLysSerLeuGlnThrThrAspThrTyr 951

DB 407 TCACTAGGTGAGAGAAAGTAAACAGAAACAGTTCATTAAGTCGAGACGAGATATG 348

QY 952 ValThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLys 971

DB 347 GTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 288

QY 972 IleThrLeuThrSerThrArgArgCysLysProThrLysCysThrThr----- 987

DB 287 GTGACGTTGACGTCGATGAGCGGAAGTAAACCTCCAAATGATACCAAGATTCAACAGA 228

QY 988 -----GlyGluGluAspAspAlaGlyAspValLysProSerGluGlyLeu 1002

DB 227 GAGACAGATTAAGAGAGAGAGAGAA-----GAGAGAAACGAAATGAGGAGATG 177

```

Oy 1003 ArgMetSerGlyTPaPnaValMetArgGlyValIleValAlaMetValIleSerPheMet 1022
Db 176 AAAATGAAGTTCCTGATATGATTAATAATGTTGTGGAGATGATTTATGGGGATG 117
Oy 1023 Ile 1023
Db 116 ATG 114

RESULT 14
AM332531/c 710 bp mRNA linear EST 31-JAN-2000
LOCUS Snp12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM332531
VERSION AM332531.1 GI:6828888
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 710)
AUTHORS Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
LOCATION/Qualifiers
1..710
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (59-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dT priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN
Alignment Scores:
Pred. No.: 5,62e-29 Length: 710
Score: 393.00 Matches: 79
Percent Similarity: 62.78% Conservative: 34
Best Local Similarity: 43.89% Mismatches: 39
Query Match: 7.26% Indels: 28
DB: 2 Gaps: 5

US-10-654-416-14 (1-1023) x AM332531 (1-710)
Oy 872 TyrValAePLeuLysGluArgCyAsnLysLeuGluSerAaPcyArgIleLeuGluAsp 891
Db 691 TATTTGGAATTCGAAGAAGAGATGCAATGCTTGAAGACTAGATGGCGTTTAAAGAGAT 632
Oy 892 CysLysAePLeuGluGluValCysLysLysIleAsnLysAlaCyAsnLysLeuLysPro 911
Db 631 TGTGAGGAGATCTAAACCAAGCTGTAAAGAAATAGAGAGATTATGCAAGAAGATAGAAC 572
Oy 912 LeuGluValLysProHisGluThr----- 919
Db 571 TTTAAAGTTGGCGCTCATCATATACAGAAATATCAAGAAATCTCAACCTTAGCGGACG 512
Oy 920 -----ValThrGluSerThrThrThrThrThrThrThrThrThrThrValAla 935
Db 511 ACTTACGGCCACTTTCACGACTACTACGACGACTACTACGACGACTACTACTACTACCAAC 452

```

```

Oy 936 AspPro-----LysAlaThr---GluCysLysSerLeuGluThrThrAspThr 950
Db 451 AAGCGGGAAGTCGAGAGAAAAGTAACAGAGAGTCAATGATACAGACCCAGATACA 392
Oy 951 TrpValThrGluThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSer 970
Db 391 TGGGTGACGAGACGCTATTCATGCAATGCAATGCAACGACTACCTACCGTACGCTG 332
Oy 971 LysIleThrLeuThrSerThrArgArgCysLysProThrLysCyThrThr-----Gly 988
Db 331 ACACTGACGTTGACTTGATGCGCAGTCAAGCTACCAATGATACTACTAGATTCAAGC 272
Oy 989 GluGluAspAaPLeuArgIle-----AspValLysProSerGluGlyLeuArg 1003
Db 271 AGAGAGACAGATTAAGAGAGAGAGAGAGAGAGAGATGTAACCAATGATGGGTGAA 212
Oy 1004 MetSerGlyTPaPnaValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db 211 ATAAGAGTTCCTGATATGATTAATAATGTTGTGGAGATGATTTATGGGATGATG 152

RESULT 15
AM334916/c 800 bp mRNA linear EST 31-JAN-2000
LOCUS S40G3 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM334916
VERSION AM334916.1 GI:6831273
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 800)
AUTHORS Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
LOCATION/Qualifiers
1..800
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (59-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dT priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN
Alignment Scores:
Pred. No.: 6.79e-29 Length: 800
Score: 393.00 Matches: 84
Percent Similarity: 47.14% Conservative: 48
Best Local Similarity: 30.00% Mismatches: 122
Query Match: 7.26% Indels: 26
DB: 2 Gaps: 3

US-10-654-416-14 (1-1023) x AM334916 (1-800)
Oy 425 GluAsnAlaCysLysAePLeuArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAla 444
Db 784 GATCAAGCGTCTCAAAATGTATGACGACGCGTCTATTAATAAGGACAAAGATAGAGTGTG 725

```



*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 29, 2005, 09:09:50 ; Search time 396 Seconds  
(without alignments)  
4227.048 Million cell updates/sec

Title: US-10-654-416-14  
Perfect score: 5411  
Sequence: 1 MARAVRQAAAGTQNSIDEEH.....MSGMNVRGIVYVAVISFMI 1023

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh  
-Q=/cg2\_1/USFTO.spool/US1065416/runat\_26102005\_101024\_21768/app\_query.fasta\_1.1159  
-DB=Issued\_Patents\_NA -OFMT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTMT=pct -NGMT=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USBR=US1065416.ccdn\_1.1.237 -runat\_26102005\_101024\_21768 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NSG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cg2\_6/pctodata/1/ina/5A.COMB.seq.\*  
2: /cg2\_6/pctodata/1/ina/5B.COMB.seq.\*  
3: /cg2\_6/pctodata/1/ina/6A.COMB.seq.\*  
4: /cg2\_6/pctodata/1/ina/6B.COMB.seq.\*  
5: /cg2\_6/pctodata/1/ina/PCTUS.COMB.seq.\*  
6: /cg2\_6/pctodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5411	100.0	3072	4	US-09-762-724-13 Sequence 11, Appl
2	4398.5	81.3	3054	4	US-09-762-724-11 Sequence 11, Appl
3	4352	80.4	3084	4	US-09-762-724-7 Sequence 7, Appl
4	3827	70.7	3090	4	US-09-762-724-5 Sequence 5, Appl
5	3401.5	62.9	3081	4	US-09-762-724-9 Sequence 9, Appl
6	2762.5	51.1	3042	4	US-09-762-724-1 Sequence 1, Appl
7	2404.5	44.4	3006	4	US-09-762-724-3 Sequence 3, Appl
8	913	16.9	2814	1	US-07-781-034-1 Sequence 1, Appl
9	913	16.9	2814	5	PCT-US92-08328-1 Sequence 1, Appl
10	559	10.3	1448	1	US-07-781-034-3 Sequence 3, Appl
11	559	10.3	1448	5	PCT-US92-08328-3 Sequence 3, Appl
12	396	7.3	249	4	US-09-762-724-15 Sequence 15, Appl

13	274	5.1	6016	4	US-09-949-016-1054	Sequence 1054, Ap
14	274	5.1	6016	4	US-09-949-016-2264	Sequence 2264, Ap
15	262	4.8	4206	4	US-09-750-590A-3	Sequence 3, Appl
16	262	4.8	4206	4	US-09-750-590A-1	Sequence 1, Appl
17	261	4.8	5574	4	US-09-917-354-40	Sequence 40, Appl
18	261	4.8	6861	4	US-09-949-016-1240	Sequence 1240, Ap
19	261	4.8	6861	4	US-09-949-016-1241	Sequence 1241, Ap
20	261	4.8	6861	4	US-09-949-016-1242	Sequence 1242, Ap
21	260.5	4.8	6773	3	US-09-166-350-27	Sequence 27, Appl
22	260	4.8	5923	4	US-09-976-594-907	Sequence 907, Appl
23	253.5	4.7	6008	4	US-09-949-016-5058	Sequence 5058, Ap
24	253	4.7	5925	4	US-09-949-016-5233	Sequence 5233, Ap
25	252	4.7	6276	4	US-09-949-016-5209	Sequence 2009, Ap
26	252	4.7	6284	4	US-09-949-016-1028	Sequence 1028, Ap
27	252	4.7	10300	4	US-09-949-016-6136	Sequence 616, App
28	251.5	4.6	5661	3	US-08-938-105-2	Sequence 2, Appl
29	251.5	4.6	7453	4	US-09-620-312D-248	Sequence 248, App
30	251.5	4.6	7501	4	US-09-620-312D-249	Sequence 249, App
31	251	4.6	8590	4	US-09-949-016-5562	Sequence 5562, Ap
32	251	4.6	38575	4	US-09-949-016-17304	Sequence 17304, A
33	251	4.6	119153	4	US-09-949-016-12378	Sequence 12378, A
34	247.5	4.6	6175	4	US-08-875-435B-1	Sequence 1, Appl
35	243.5	4.5	7596	4	US-09-023-655-1463	Sequence 1463, Ap
36	241	4.5	6644	4	US-08-875-435B-5	Sequence 5, Appl
37	240.5	4.4	6011	4	US-09-875-435B-1017	Sequence 3017, Ap
38	237.5	4.4	10136	5	US-08-353-700-2	Sequence 2, Appl
39	237.5	4.4	10136	5	PCT-US95-16216-2	Sequence 2, Appl
40	237	4.4	9626	4	US-09-150-867-2	Sequence 2, Appl
41	232	4.3	8789	1	US-08-328-254-5	Sequence 5, Appl
42	231	4.3	4868	1	US-08-139-937-12	Sequence 12, Appl
43	231	4.3	4868	5	PCT-US93-11310-12	Sequence 12, Appl
44	231	4.3	5883	4	US-09-949-016-5001	Sequence 5001, Ap
45	228	4.2	4407	4	US-09-949-016-1690	Sequence 1690, Ap

#### ALIGNMENTS

RESULT 1  
US-09-762-724-13  
; Sequence 13, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; CURRENT FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; PRIOR FILING DATE: 1998-08-17  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 3072  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3072)  
US-09-762-724-13  
Alignment Scores:  
Pred. No.: 0  
Score: 5411.00 Length: 3072  
Percent Similarity: 100.00% Matches: 1023  
Best Local Similarity: 100.00% Conservative: 0  
Query Match: 100.00% Mismatches: 0  
DB: Gaps: 0  
US-10-654-416-14 (1-1023) x US-09-762-724-13 (1-3072)

1 MetAlaATGAlaVallysaTgInAlaAaglyThrGlnaenSerTleAspGluGlnHis 20  
1 ATGGCGCGGGGCGTCAAGCGGACAGCAGAGACACAGAAATGAGATTGATGAGAGAACAT 60  
21 ValLeuAlaLeuIleLeuLeuGlnaAspGlyLeuSerGluGlnGluCysValyLeuLeu 40  
61 GTTTTAGCTTTAAATTTCTAAAGGAAAGATGACTAACTGACAGAAATGCAAAAAAACTTA 120  
41 LybLybTyrcYsGInGluLeuThrcGlnAlaLybLeuAenIleGluGlnValHisArglys 60  
121 AAAAAAATTTGCCAAGAATTCGTCGAACCAAACTTAATATAGAACAAAGTACACAGAAA 180  
61 LeuLybGlyPheCysGluuAspGlyLybAlaAspThrLybCysLybGluLeuLybAlaAsn 80  
181 CTTAAAGGTTTTTGGGAGATGAGAAACCAAGATCAAAATGCAAAAGAACTGAAAGCCCAAT 240  
81 IleGluLybLybCysbThrThrIleLybGlyLybLeuLybGluAlaIleLybLybLybIle 100  
241 ATTGAGAAAAATGTACTTAACAATCAAAAGAACTTAAGAAAGCAATTTAAAAAAAATTT 300  
101 GlnIleIleThraAspLybAspCysLybGluuAspGluuAspGlnGlnCysLeuPheLeuGluGly 120  
301 CAGATTATTAAGGATTAAGATTGGCAAGAGATGAAACAACATGCTATTTTGGAGGGA 360  
121 ValCysSerLybGluLeuLybAspAspCysaenThrLeuAtgAenLybCysTyrgLnyls 140  
361 GTATGTTCAAAAGAACTTAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420  
141 LybAspAspLybValAlaGluGlnValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160  
421 AAAGGCTAAAGATTGCGGAAAGAGTCTTTTAAAGACATTCGTAGGACATTTAAATGGA 480  
161 SerValIleCysGluLybLybLeuLybGluIleCysProValMetGlyArgGluSerAsp 180  
481 TCAGTCATATGTGAAAAAAAATTTAAAGATTTCCTGTCATCGGAGGAGAAAGTAT 540  
181 GluLeuThraenLeuCysLeuAsnGlnLybGluThrcyLybAsnIleLeuIleGluLyb 200  
541 GAGTTAAACAAACTGTGTCTGAACCAAGAAAGACATGTAAGATTTTAAATGAAAAA 600  
201 AspLybLybCysGlyThrLeuLybThraAspValSerAlaIleGluGlySerPheLybLyb 220  
601 GATTAAGAGTCCGGTACTTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAAA 660  
221 GluThrcyLeuGluLeuLeuGlnGlnCysTyrcPheTyriIleGlyAsnCysGlyAspAsp 240  
661 GAAACATGCTTGAATTACTCGAACAAATGCTATTTTACATTGGAAATTCGCGAGAGACAC 720  
241 AspIleIleLybCysIleGluLeuGlyGlyLybCysGlnGluGlnAenIleAlaTyrcMet 260  
721 GATTAATTAATTAATGTAATTTGAAATGGGAGAAATGCCCAAGAAACAATAATGCTTATATG 780  
261 ProProGlyProAspPheAspProThraAspProGluAlaThrIleAlaGluAspIleGly 280  
781 CCACCGAGACCCGATTTTGAATCCACTAGCCAGGCTTACAAATAGCAGAGCATTAAGG 840  
281 LeuGluGluPheTyrcLybLybValGluGluAspGlyValPheIleGlyLybAsnHisLeu 300  
841 CTGGAGAGGTTTTTAAGAAGTAGAGAGATGAGTTTTTATTTGAAAAGATCATCTA 900  
301 ArgAspAlaThraIleLeuLeuAlaLeuLeuIleGlnaAspSerSerLeuLybLybAsp 320  
901 AGAGATGCGACAGCTTTGTTGGCATTTGTGATCCCAAGATTCATAGCTTTAAAAAAAAC 960  
321 AspLybGlyLybCysGluGluAlaLeuGlnLybSerCysLybAsnProHisGluHisGlu 340  
961 GACAAAGGAAATGCGAAGAGCCCTTCAAAAGACCTCAAAATCCCTCAAGAACATGAG 1020  
341 AlaLeuGluSerLeuCysLybLybAsnGlyLeuSerAsnAspGlyThrLybLybCysGly 360  
1021 GCTTTAGAAAAGTTTATGTAAAGAAAAATGCTTTAACTTAATGATGGAACGAAAAAATGTGAA 1080  
361 GluLeuGlnaAspIleAsnLybThrcyLybIlePheThrSerLybValThraAsnAsn 380

1081 GAATTGCAAAAATGATTATTAACAAAATTTGCAAAATTTTCACTTCAAAAGTACATTAAT 1140  
381 ArgLeuPheAspProThrLybGlyAsnAsnGluIleValGlyTyrgGluGlyLeuProThr 400  
1141 CGTCTTTTATGATCCAAAGAAAGGAAATTAATGAAATTTGTTGATGCGAAAGGTTCCAAACA 1200  
401 PheLeuSerAsnGluuAspCysAlaLybLeuGluuSerTyrcCysPheTyrcPheGlnLybLyb 420  
1201 TTTCTTAGCAACGAAGATTGGCGCAAAATTTGAGTCCATTTGTTTCTATTTTGAAAAAAA 1260  
421 CysProAspGlyGluAenAlaCysLybAsnIleArgAlaThrcyTyrcLybAspGlyLeu 440  
1261 TGTCAGATGAGAAAAATGCAATGTAATAAATTAAGACCAACATGTTACAAAGAGGACTT 1320  
441 AspAlaATGAlaAenLybValLeuGlnGluuAsnMetArgGlyMetLeuHisGlySerAsn 460  
1321 GATTCAGCGGCAAAATTAAGTCTGCAAGAAATTTGCGAGGAAATTTTACATGCTTCAAAAC 1380  
461 LybSerTrpLeuGluLybPheGlnGlnGluLeuValLybValCysGluLybLeuLybLyb 480  
1381 AAAAGCTGGCTTGAAAGTTTCAACAGAAATTAAGTAAAGATGATGCAAGAACTGAAAAA 1440  
481 GluAsnLybGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLyb 500  
1441 GAAAAACAAGAGAGTTTCTCAAAAGATGAATTAATTTATCTGTGTGACAGCCAGCAAAA 1500  
501 AlaAlaATgLeuLeuThraAspLeuArgMetLybThrIlePheLeuAspGlnGlnLeu 520  
1501 GCAGCCGGCTTGCTTACATCATGATCTTCAATGAAACTATCTTTTACGACAAACAACCTG 1560  
521 AspGlnLybAspAspPheProThraAspLybAsnCysLybGluLeuGlyTyrcGln 540  
1561 GATCAAAAGCGAGATTTCCGCAAGATTAATAAATTTGCAAGAAATTTGGGAGAAATGGCCAA 1620  
541 AspLeuGlyGluAspSerLybGluIleThrTrpProCysHisThrLeuGluGlnGlyCys 560  
1621 GATTTAAGAGAGATTCATAAAGAAATTAACATGATGATCAATGACATGAGACACAAATGC 1680  
561 AsnArgLeuGlyThrThrGlnIleLeuLybGlnValLeuLeuAspGluHisLybAspThr 580  
1681 AATGCTTTGGGACTACAGAAATTTTAAGCAGCTTTTATTTGATGAACAACAAGATACT 1740  
581 LeuLybAspGlnGluSerCysValLybTyrcLeuLybGluLybCysAsnLybTrpSerArg 600  
1741 TTGAAAGACCAAGAAATTTGTGTAATAATACCTTAATAAAGATGTAATTAATGTTCTGA 1800  
601 ArgGlyAspAspAspPheSerPheValCysValPheGlnAsnAlaThrcyGluLeuMet 620  
1801 AGAGGAGATGACCGTTTCTTTTGTATGTCTTCCAAAACGCTAGCGTGAAGCTGATG 1860  
621 ValLybAspValLybAspAspArgGluValPheLybLybAsnIleLybAlaSerTyrcIle 640  
1861 GTAAAGACGTAAGACAGGTGTGAAGTATTCAAAAAAATTAATAAGCTTCATATATT 1920  
641 IleGluPheLeuGlnaAenThraAsnLybIleThrThrLeuGluuArgAsnCysProSer 660  
1921 ATTGAATTTCTTGAATAATTAATACAAATTAATAACAACACTGGAAGAAATTTGCCCTCT 1980  
661 TrpHisThrTyrcCysaenArgPheSerProAsnCybProGlyLeuThrLybGluuAspSer 680  
1981 TGGCATGCGATGCAATGATGTTTTTCACTTAATGTCCAGGTCTTACGAAAGAAATAGT 2040  
681 CysThrLybIleLybLybHisArgGluProPheTyrcLybAspGlybAlaLeuGluAspAla 700  
2041 TGTAACAATAATCAAGAGCATGTAAGCCGTTCTTAATAAGAAAGGCTTCGAAGATGCT 2100  
701 LeuLybValGluLeuGlnGlyLybLeuThraAspLybSerLybCysGluProAlaLeuLyb 720  
2101 CTCAAGTAAGAGCTTCAAGGAAATTTGATGAATAATCTAAATGTAACCTGCAATTTGAAA 2160  
721 ArgTyrcCysThraValAlaGlyAenValAsnAsnAlaSerIleSerGlyLeuCysLybAla 740



```

Db 2161 AGATATTGTACAGTAGCGGAAACGTAAATATGCGTCMAATCGCTTAATGCAAACT 2220
Qy AanthrilysaerpaenserGlylyseSerapGluabepalaaglysgluLeuCyegluys 760
Db 2221 AACCCAGAGTAATCTCTGGAAAGAGTATAGATGCTTAAGAAAGAACTCTTGAGAAA 2280
Qy LeuVallysgluValgluGluGlylysaValleuprothrgluLeuGlygluProala 780
Db 2281 TTACTGAAAGAGTAGAAGAACAGTGCAGAACATTACCAACAGATTATGACCAACCGGCA 2340
Qy 781 AlalapleuVallysaerpylylyethrlyrlygluGluLeuVallysaValleuGluVala 800
Db 2341 GCTGATCTAAAAAAGATTATAGACATATAGAACTTAAGAAACGTCGAGAGAAAGCA 2400
Qy 801 MecanlysaerSerleuValleuSerleuVallylysaValleuSerleuVallysaer 820
Db 2401 ATGAAACAGTCACTCTTTGTTGCTCACTCACTTAAGAAAGCAAGTAATGATCAAAA 2460
Qy 821 Seranberlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 840
Db 2461 AGTAATGCAAAAACAGGATAGATGCGTTCAACGAGATTCAAGATACCAACAAA 2520
Qy 841 HlaVallylyleuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 860
Db 2521 CATGTGAAAAATCTACGAGGAGGAGTTAAGGATGATCCGTAAACAGATTAGAAAGCTAAA 2580
Qy 861 AlalapleuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 880
Db 2581 GCATTGATTTGGGACGAGGAGTTGGAAGATGATGATGATGATGATGATGATGATGATGATGAT 2640
Qy 881 LyseuGlyseuSerapCyarsglylyleuGlylysaerlysaerlysaerlysaerlysaerlysaer 900
Db 2641 AAATTTGAGATCAGATTGAGAAATTAAGAGAGATTGCAAAAGACTTAAGAAAGATGCAAA 2700
Qy 901 LyseuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 920
Db 2701 AAGATTATTAAGGCTTGTGCAATCTGAAGCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGT 2760
Qy 921 ThrGlyserThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 940
Db 2761 ACAGAAAGTACAGACACACACACACACACACACACACACACACACACACACACACACACACAC 2820
Qy 941 GluCylyseSerleuGluThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 960
Db 2821 GAATGCAAAATCTTACAGACACACACACACACACACACACACACACACACACACACACACAC 2880
Qy 961 ThreSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 980
Db 2881 ACCTCTACTATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2940
Qy 981 Lyseprothlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 1000
Db 2941 AAACCAACCAAGTACACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 3000
Qy 1001 GlyleuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 1020
Db 3001 GCGGTGAGAGATGAGTGGTGAATGTATGAGGGGGGAGATGATGATGATGATGATGATGATGAT 3060
Qy 1021 Pheneticle 1023
Db 3061 TTCATGATT 3069

```

```

; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096, 805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3054
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; NAME/KEY: CDS
; LOCATION: (1) . (3054)
US-09-762-724-11

Alignment Scores:
Pred. No: 0 Length: 3054
Score: 4398.50 Matches: 838
Percent Similarity: 88.81% Conservative: 75
Best Local Similarity: 81.52% Mismatches: 98
Query Match: 81.29% Indels: 17
Gaps: 9

US-10-654-416-14 (1-1023) x US-09-762-724-11 (1-3054)

Qy 2 AlapgaVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 20
Db 1 GCGGGGCGGTCACAGCGGACGTAACAGACATCAAGGCAATGATGATGATGATGATGATGATGAT 60
Qy 21 ValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 40
Db 61 ATTTGGCGTTGATTCTTACAAAGAGATGCAATGAGATGCAAAATGCAAAATGCAAAATGCAAAATGCA 120
Qy 41 LyseuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 60
Db 121 GAAATAATCTGCGAAGGTTGAAAAAGCATCACTAACATGAAAAAGTACATTAATG 180
Qy 61 LeuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 77
Db 181 CTTAAAGATTTCTGTGAAATGGAAGCAAGTAACCAATACAAATGCAAGTCTA 240
Qy 78 LyseuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 96
Db 241 CAAGCCAAAGTTACGGGGAATGTACAAATTTTAAACAAAGGATGAGACCAAGCTTA 300
Qy 97 LyseuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 116
Db 301 ACAAAATCA-----TCAGATGATTAATGCAAAAGAGTGAACGACAAATGCTTA 348
Qy 117 PheluGlylysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 136
Db 349 TTTTGGAGGAGCATC---CATATCTTGTACAGATGTTAACTTAACCAATGATGATGATGATGATGAT 405
Qy 137 CyseuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 156
Db 406 TGTTACAGAAAAAAGCTGACGAGTAGAGAAAGTCTTTGAGGCGCCTTCTAGT 465
Qy 157 AspleuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 176
Db 466 GATCTCAATTAACAGAAACATGAAAAAACTGAAAAAGATTTCCAGCTTTCCAG 525
Qy 177 ArgGlyserapGlyleuThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 196
Db 526 AGGAAAGTAAATTAATTAACGACTTGTGTTGAACCAAGAAAAAGCTGCGAATATT 585
Qy 197 LeuVallysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 216
Db 586 ATAAAGAAAAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 645
Qy 217 SerThelysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaerlysaer 236
Db 646 AGTTTAAAAAAGAAATATGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 705

```

```

RESULT 2
US-09-762-724-11
; Sequence 11, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762.724
; CURRENT FILING DATE: 2001-02-09

```

[illegible]

Db	1774	AAATAATGGCTTAGAAGAGAGATGACCGCTTTCTCTTTTGTATGTGCTTCCAAAACGCT	1833
Oy	616	ThrCysGluLeuWecValIysAspValIysAspArgGluValPheIleYasAsnIle	635
Db	1834	ACGTGTGAGCTGATGGTATAAAGACGTGAAGACAGGTGTGAAGTATTCAAAAAAATATTA	1893
Oy	636	LysAlaSerTyrIleIleGluPheLeuGluAAsnAsnThrAsnLysIleThrThleuGlu	655
Db	1894	AAAGCTTCATATATATTAATTTCTGTGAAAATATATACAAATATAACAACTGGAA	1953
Oy	656	ArgAsnCysProSerTyrpIsthrTyrCysAsnArgPheSerProAsnCysProGlyLeu	675
Db	1954	AGAAATGTGCTCCCTTGGCATACGATTTGGCATATGATTTTCACCTAATTTGTCAGGCTT	2013
Oy	676	ThrIysGluAsnSerCysThrIlyslleYlyshIAsrGluProPheTyrIlyAsrIlyAs	695
Db	2014	ACGAAAGAGATAGTGTGACAAAATTCAGAAAGCATTTGAGAGCGGTCTATATAAAGAAAG	2073
Oy	696	AlaLeuGluAspAlaLeuLysValGluLeuGlnGlyIlyLeuThrAspLysSerIlyCys	715
Db	2074	GCGTTGGAAGATGCTCTCAAGTAGCGCTTCAGGAAAAATTTGACGATTAATCTTAAATGT	2133
Oy	716	GluProAlaLeuLysAsrGlyrCysThrValAAsIyAsnValAAsnAlaSerIleSer	735
Db	2134	GAACCTGCATTGAAAAGATTTGTACAGTACGGGAAACGTAAATATATGCGTCAATCGT	2193
Oy	736	GlyLeuCybelysAlaAsnThrIysAspAsnSerGlyLysSerAspGluAspAlaArgLys	755
Db	2194	GGCTTATGCAACACTGACCAACCAAGATTACTGTGAAAGAGATGAGATGCTAGAAAG	2253
Oy	756	GluLeuCybelysLeuValIysegIuValGluGluGlnCybelysAlaLeuProThrGlu	775
Db	2254	GAACCTGCTGAGAAATTTAGTGAAGAGTGAAGAACGTGCAAGACATTCACCAACAA	2313
Oy	776	LeuGlyGluProAlaAlaAspLeuLysAspTyrIlyserThrTyrGluGluLeuLysAs	795
Db	2314	TTTAGGCAACCGGCACGCTGATTTAAAAAAAGATTTATTAAGACATTAAGAGAACTTAAAGAA	2373
Oy	796	ArgAlaGluGluValaIeAsnLysSerSerLeuValLeuSerLeuIleLysIysAsnGlu	815
Db	2374	CGTGCAGAGGAACATGACAAAGTCACGCTGTGTTGTCTACATTTAAGAAAAACGAA	2433
Oy	816	SerAsnValSerIysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeu	835
Db	2434	AGTAAATGATCAAAAAGTAAATAGCAAAAACAAAGATTAAGATGCCGTTTCAACCGACCTT	2493
Oy	836	GluAspThrThrIyshiIeValIyshiIeLeuArgGlyValIyAspValSerAlaThr	855
Db	2494	CAAGATACCACAAAACATGTGAAAAATCTACAGAGAGAGGTTAAGAGATGTACCTTAACA	2553
Oy	856	GluLeuGluValaIyAspAlaPheAspLeuAlaGluValPheGlyArgTyrValaAspLeu	875
Db	2554	GAATTAGAACCTTAAGCAATTTGATTTGGCAGCAGAAAGTATTTGGAAGATATGATATTG	2613
Oy	876	LysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeu	895
Db	2614	AAAGAAAGATGTATAATTTGGATCAATTCAGAAATTAAGAGAGATTCGCAAAAGACTTA	2673
Oy	896	GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys	915
Db	2674	GAAGAAGTATGCAAAAAGATTAATAGGCTTGTCCAACTGCAACCTCTGAGGTGAAG	2733
Oy	916	ProHisGluThrValaIyGluSerThrThrThrThrThrThrThrThrThrThrValaIa	935
Db	2734	CCGACGAAACAGTGAAGAAAGTACAGACATCAACAAACAAACAAACGCTGCC	2793
Oy	936	AspProLysAlaThrGluCysLysSerLeuGlnThrThrAspThrTyrValaThrGlnThr	955
Db	2794	GATCCGAAGCAACGGAATGCAAAATCTTACAGACAAACAGACACATGGGTTTACACAGCA	2853
Oy	956	SetThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrIeuthr	975

Db 2854 TCACACACACAGCAGCTCTACTATCACTCTACATCATCAATCAATCAATGACA 2913  
Qy 976 SerThrArgArgCysGlyProThrThyCysThrThrGlyGluAspAlaGlyAsp 995  
Db 2914 TCACGAGGCGATGCAACCAACCAAGTGAAGCAGG-----GATGATGCAAGAAC 2967  
Qy 996 ValIysProSerGluGlyLeuArgMetSerGlyTyrAsnValMetArgGlyValIleVal 1015  
Db 2968 GTAGAGCAATGAGAGCTTGAAGGCTGACGCGGTGGAATGTGAATGAGCGGCGTATGTA 3027  
Qy 1016 AlameValIleSerPheMetIle 1023  
Db 3028 GCAATGCTTATTTGTCATGATT 3051

RESULT 3  
US-09-762-724-7  
Sequence 7, Application US/09762724  
Patent No. 6664053  
GENERAL INFORMATION:  
APPLICANT: Kovacs, et al.  
TITLE OF INVENTION: Identification of a region of the major surface  
glycoprotein (MSG) gene of human Pneumocystis carinii  
FILE REFERENCE: 4239-58054  
CURRENT APPLICATION NUMBER: US/09/762,724  
CURRENT FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: PCT/US99/18750  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: US 60/096,805  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 3084  
TYPE: DNA  
ORGANISM: Pneumocystis carinii sp. f. hominis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3084)  
US-09-762-724-7

Alignment Scores:  
Pred. No.: 0 Length: 3084  
Score: 4352.00 Matches: 826  
Percent Similarity: 87.28% Conservative: 80  
Best Local Similarity: 79.58% Mismatches: 106  
Query Match: 80.43% Indels: 26  
Gaps: 6

US-10-654-416-14 (1-1023) x US-09-762-724-7 (1-3084)

Qy 1 MetAlaArgAlaValIleValArgGlnAlaIleGlyThrGlnAsnSerIleAspGluGluHis 20  
Db 1 ATGCGCGGCGGCGTCAAGCGGCGGCAAAAGTGCACAGATACCATTTGATGAGAGACAT 60  
Qy 21 ValIleuAlaLeuIleLeuIleGluAspGlyLeuSerGluGluGlnGluCysGlyLeuLeu 40  
Db 61 GTTTTACCTTTGATTTTAAATAAATGATGATTAGAGATTACAAATGCAAAATCTAAGTTG 120  
Qy 41 LysLeuSerGlyCysGlnGluLeuThrGluAlaLeuValLeuAsnIleGluGlnValHisArgLys 60  
Db 121 GAAGAATATTGCAAAACATTAAACAAATGCAAGATTAAATCCAGAAAAAGTTCAAGAAAA 180  
Qy 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysGlyGluLeuLysAlaAsn 80  
Db 181 TTAAAGATTTCTGTGATTAACGGAGAAACGAATGAAATGCAAAATCTTAAAGAACAA 240  
Qy 81 IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIle 100  
Db 241 GTCAATCAAAATGCAATTAATTTCAAGGAAATCTTCAACAGCTGCTCAAGAAAAATTT 300  
Qy 101 GlnIleIleThrAspLysAspCysGlyGluAsnGluGlnGlnCysLeuPheLeuGluGly 120  
Db 301 TCAGAAATTAACAGATGAGATTGCAAAAAAGAAATGAAACAAATCCATATTTTTCGAGGA 360

Qy 121 ValCysSerLysGluLeuLysAspAspCysAsnThrIleuArgAsnLysCysTyrGlnLys 140  
Db 361 GCATGTCCACAGAACTTAAGATGACTGCATTAATTAAGAAATTAATCTTTACAAAA 420  
Qy 141 LysArgAspLysValIleGluGluValIleLeuArgAlaLeuArgSerAspLeuAsnGly 160  
Db 421 GAACGGAACAATGGCAGAAAGATTCTTTGAGGCGCTTGTGTGATCTCAATGA 480  
Qy 161 SerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGluSerAsp 180  
Db 481 ACAAGACATGTAAAAAATGAAAGAGTTGCCGCAAAATTAAGAAAGAGGAT 540  
Qy 181 GluLeuThrAsnLeuCysLeuAsnGluLysGluIleCysPheAsnIleLeuIleGluLys 200  
Db 541 GAATTAACGAGAGCTTTGCTTTTCAAAAAACAATGCTTAAGTCTTGTAAACAAAA 600  
Qy 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaLeuGlySerPheLys-- 219  
Db 601 AAAAGTAAATGTATCTTGAAGAAAGATTGAAGAACCTTAAGAAATGAATG 660  
Qy 220 LysGluThrCysLeuGluLeuLeuGlnCysTyrPheTyrIleGlyAsnCysGlyAsp 239  
Db 661 CGAAGAAATGTCTACTATTACTTGAAGCATGTACTTTCACAGAGGAACTGAAAGA 720  
Qy 240 AspAsp-----IleIleLysCys 245  
Db 721 GACAAATCAAGTCAATTAACCTTAATTAAGACTGCAGAAAGATTTGACAGAGT 780  
Qy 246 IleGluLeuGlyLysCysGlnGluGlnAsnIleAlaTyrMetProGlyProAsp 265  
Db 781 GATGAATTAACAGAAAGTGTGAAAAAATATTTGTTATATGATCCAGATCCGAT 840  
Qy 266 PheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyr 285  
Db 841 TTGATTCACCACTAAGCAGAGCCTACACTGACAGAGACATAGGCTGAGAGCTTAT 900  
Qy 286 LysLeuValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAla 305  
Db 901 AAGAGGCGAGAGAGATGGAATTTTGTGTAAGACACATGTAAAGATGCAACAGCT 960  
Qy 306 LeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysAspAspLysGluLysCys 325  
Db 961 TTGTGGCACTACT-----CTTAAGAAACCTTAAAGAAAGAAATGT 1005  
Qy 326 GluGluAlaLeuGlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluSerLeu 345  
Db 1006 ATAAAGCCCTTAAAAAACTCGAAACCTCATGAAATGATGAGGCTTAAAGAAATCTA 1065  
Qy 346 CysLeuLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAsp 365  
Db 1066 TGTAAAGAAATTAACCAAGTATGATGGAACCAAAAAATGATGAATCAAAAAAGAT 1125  
Qy 366 IleAsnLysThrCysLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspPro 385  
Db 1126 GTTAACAAAATCTTCAACATCTTCAATCAACATCTTAAAAACGCTTTTCAATTCA 1185  
Qy 386 ThrLysGlyAsnAsnGluIleValGlyTyrGluGlyLeuProThrPheLeuSerAsnGlu 405  
Db 1186 CTTGATGGA-----ATTGCGGAATGGGGAATTAACCGCATTTCTTAAGATGA 1236  
Qy 406 AspCysAlaLysLeuGluSerTyrCysPheTyrPheGluLysCysProAspGlyGlu 425  
Db 1237 GATTGTCAAAATCAAGATCTTATGCTTTTATTAATAAAGAACTGTCCAGATGTCAA 1296  
Qy 426 AsnAlaCysLysAsnIleArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsn 445  
Db 1297 GAGCTTGATGATGATGAGGAGGAGCTTTTACAGAGAGGCTTATGACGCGGCAAC 1356  
Qy 446 LysValLeuGlnGluAsnMetArgLysMetLeuHisGlySerAsnLysSerTyrPheGlu 465  
Db 1357 AGTGTGTTGCAAAAAATATAGCGTGGGTATTTACGTGTTCAATCAAAAGTGGCTTAAG 1416

466 LysPheGInGInGluLeuValIysValCySGluYbLysGluYbAsnLysGlySer 485  
1417 GAGTTTCAACAAAGATGATAAGTATGTAAGGCTA---AAAGAAATAAAGAGAT 1473  
486 PheSerAAsnAPGluLeuPheIleuCybValGInProAlaIysAlaIArgLeu 505  
1474 TTCCAAACGATGAATATTTCTGTGTGTACGCGCAAAAGCTGCGAGATTACTT 1533  
506 ThrHISAPLeuArgMetIleYThrIlePheLeuArgGInGInLeuAPGluYbArgAsp 525  
1534 ACACACGATCATCAATAGAGGCTACCTTTTACGACAAACAATTGATCAAAAGAGAGAT 1593  
526 PheProThraAPLeuYbAsnCybYbGluYbGluYbArgYbCybGInAPLeuGlyGluAP 545  
1594 TTTCCGACAGATTAAGACTGCAAGGAGACTGAGGAAAAATGCAAGATTGAGAAAGAT 1653  
546 SerIySGluIleThrTriProCybHISThrLeuGInGInCybAsnArgLeuGlyThr 565  
1654 TCAGAGAAATTAATCACTGGCCATGTCTATCACTGGAGCAATGCAATCGCTTGGGACT 1713  
566 ThrGluIleLeuYbGInValLeuLeuAPGluHISLysAPThrLeuYbArgGInGlu 585  
1714 ACAGAAATTTTAAACAGGTTTATTGATGTAACAAGATATTGTAAGACCAAGAA 1773  
586 SerCybValIySTyTrLeuYbGluYbCybAsnLysSTPserArgArgGlyAPAsnArg 605  
1774 AGTTGTGTAATATCACTTAAGAAAGATGTAATTAATGCTTGAAGAGAGATACCGT 1833  
606 PheSerPheValCybValPheGInAsnAlaThrCybGluLeuMetValIysAPValIys 625  
1834 TTCTCTTTGTATGTCTTTTCCAAACGCTACGTGAGCTGATGTAAAGACCTGAAA 1893  
626 APAPArgCybGluValPheIyLysAsnIleYbAlaSerTyTrIleGluPheLeuGlu 645  
1894 GACAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATATATTGAATTTCTTGA 1953  
646 AsnAsnThraAsnLysIleThrTrLeuGInArgAsnCybProSerTriHISThrTyCyb 665  
1954 AATAATTAACAAATAAATAACAACACTGGAAGAAATTTGCTCTTGACATACGATATTC 2013  
666 AsnArgPheSerProAsnCybProGlyLeuThrLysGluAsnSerCybThrLysIleYb 685  
2014 AATATATTTTCACTTAATTTGTCACAGCTTACGAAAGAAATAGTTGACAAAATCAAG 2073  
686 LysHISArgGluProPheTyTrIyArgIyAlaLeuGluAPAlaLeuYbAlGluLeu 705  
2074 AAGCATGTGTAGCCCTCTATTAAGAAAGCCCTTGGAAGATGCTCTCAAAAGTAGACTT 2133  
706 GInGlyYbLeuThraAPLysSerIySGluProAlaLeuYbArgTyTrCybThrVal 725  
2134 CAAGGAAATTAATGATTAATCTTAATGTGAACCTGCAATGGAACAGATTTGTACAGTA 2193  
726 AlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCybYbAlaAsnThraYbAsn 745  
2194 GCGGGAACGTAAATAATCCGTCATCACTGCTTATTCGAAAGCTAACCAAGAGATAC 2253  
746 SerGlyYbSerAsnGluAPAlaArgIySGluLeuCybGluYbLeuValIySGluVal 765  
2254 TCTGGAAGAGATGAGAGATGCTAGAAAGAACTCTGTGAAATTCAGTGAAGAAAGTG 2313  
766 GInGluGInCybYbAlaLeuProThrGInGluGInProAlaAlaAPLeuYbLys 785  
2314 GAAAGAACGTGCAAAAGCATTAACAAAGATTAAGAACCAAGCGAGCTGATTAAGAAA 2373  
786 APPTyTrLysThraGlyGluGluLeuYbArgAlaGluAlaMetAsnLysSerSer 805  
2374 GATTATTAAGACATTAAGAGAACTTAAGAACTGAGAGAAAGCAATGAACAAGTCCAGT 2433  
806 LeuValIleuSerLeuIleYbLysArgInuSerAsnValSerIySerAsnSerIyAsn 825  
2434 CTGTGTTTGTCACTCATTAAGAAAAACGAAGTATGTATCAAAAAGTATATGCAAAAAC 2493  
826 LysAPLeuYbAsnAlaValSerAsnGlyLeuGInAPThrThraYbHISValIleLeu 845

2494 AAGGATTAAGATGCGTTTCAACGAGCTTCAGATCCACAAAACATGTGAAATTACTA 2553  
846 ArgArgGlyValIyAsnValSerValThrGluLeuGluAlaIyAlaIAsnLeuAla 865  
2554 CGGAGAGAGATTAAAGATGTATCCGTAAACGAATTAAGAGCTTAAGCATTTGATTGGCA 2613  
866 AlaGluValPheGlyArgTyTrValAPLeuYbGluArgCybAsnLysLeuGluSerAsp 885  
2614 GCAGAGATATTGGAGATATGTAGATTGGAAGAAAGATGTAAATTAATGGAAATCAGAT 2673  
886 CybArgIleYbSGluAPCybYbAsnLeuGluGluValCybLysYbIleAsnLysAla 905  
2674 TCGCAATTAAGAGGATTTGCAAGACTTACAGAAAGTATGCAAAAGATTAAAGGCT 2733  
906 CybArgAsnLysYbProLeuGluValIyProHISGluThraValIthrgLysSerThr 925  
2734 TGTGCAATGTGAAGCCTCGAGAGTGAAGCCGACGAACAGTACAGAGATCAACAG 2793  
926 ThrThrThrThrThrThrThraValAlaAPProLysAlaThrgLysSerLeu 945  
2794 ACAACTACAAACAAACAAACAAACCGTTGCCGATCCGAAGGCAACGGAATGCAATCCCTTA 2853  
946 GlnThraAPThrAPThrTPValIthrgInThrSerThraHISThrSerThrIleThr 965  
2854 CAGCAACAGACATGGGTACACAGACATCGACACACACACACAGCTTACTATCA 2913  
966 SerThraIleThrSerLysIleThrLeuThrSerThraArgCybYbProThrLysCyb 985  
2914 TCTACATCAATCAAAATAATCAATTGACATCAACGAGGCGATCAAAACCAAGCT 2973  
986 ThrThrglyGluGluAPAPAlaGlyAPValIyProSerGlyGluYbArgMetSer 1005  
2974 ACAGACAGG-----GATGATGCGAAGACGTGAAGCAACGTGAAGGCTGAGGGTGC 3027  
1006 GlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023  
3028 GGGTGGATGTGATGAGGGGGGTATGTACCAATGTTATTTGTTTCATCATATT 3081

RESULT 4  
US-09-762-724-5  
: Sequence 5, Application US/09762724  
: Patent No. 6664053  
: GENERAL INFORMATION:  
: APPLICANT: Kovacs, et al.  
: TITLE OF INVENTION: Identification of a region of the major surface  
: FILE REFERENCE: 4239-58054  
: CURRENT APPLICATION NUMBER: US/09/762,724  
: PRIOR FILING DATE: 2001-02-09  
: PRIOR APPLICATION NUMBER: PCT/US99/18750  
: PRIOR FILING DATE: 1999-08-17  
: PRIOR FILING DATE: 1998-08-17  
: NUMBER OF SEQ ID NOS: 26  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 5  
: LENGTH: 3090  
: TYPE: DNA  
: ORGANISM: Pneumocystis carinii sp. f. hominis  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (1)..(3090)  
US-09-762-724-5

Alignment Scores:  
Pred. No.: 0 Length: 3090  
Score: 3827.00 Matches: 728  
Percent Similarity: 81.54% Conservative: 120  
Best Local Similarity: 70.00% Mismatches: 164  
Query Match: 70.73% Indels: 28  
DB: 4 Gaps: 8







Db	655	AAGAAAGATGAGCTTACGAGGAAATGTTTACACCTGCTGAACGATGTATTTTATAGA	714
Qy	235	GIvAnCyGgiYaerPaerPaRiIeIleYCyviiIegu-----	247
Db	715	GGGAATTGT-----GAAGATATATCAAAATGTAAATATCATCCGAAGACTGTATAGA	768
Qy	248	-----LeuGIyIuYvCyvGInGIuInAnIleAlTYmeC	260
Db	769	TATTTGCCAGTGTGTATACATTGGCAGGTGAATGTGAAGAAATTAAGATTATTTATACA	828
Qy	261	ProProGIYProaerPhearPhearProthrxarProGIuaIaThriIeAlaGIuAerIleGIY	280
Db	829	CATCCGGGATCCGATTTCAATTCCAACTAAGTCMAAGCCTACTGTACGAGAAACATACGA	888
Qy	281	LeuGIuGIuPheTYrLYvLYvValGIuGIuAerGIYAlPheIleGIYLYvAAnIleu	300
Db	889	CTGGAAGAGCTTTATTAATAAGCCGACAGAAAGAGGTGTTCATATTGGAAAGCTTCGTGA	948
Qy	301	ArgaerAlaThrAlaLeuLeuAlaLeuLeuIleGIaerPaerSerIeYLYvLYvAerP	320
Db	949	AGAAATGCMACTGCTCTACGTGGCGCTTTTGATTCAAAATCTAAGATCCTTAAGACTCAAGTG	1008
Qy	321	AerPYvGIuLYvCYvGIuGIuAlaLeuGIuInLYvSerCYvLYvAAnProHIeGIuHIeGIu	340
Db	1009	GCTTAAGAA---TGGCAAAAGCTTTTAAGATATCTTAAGAGTTTAAAAAGTCATGAA	1065
Qy	341	AlaLeuGIuSerLeuCYvLYvLYvAAnGIYLeuSerAerPaerGIYThrLYvLYvCYvGIu	360
Db	1066	ATTTTGGGAGATTTTGTATATCAAAATGTACTGTCTCAAAATGAATTAAGAAAGGTATAA	1125
Qy	361	GIuLeuGIuAerPaerIleAerNLYvThrCYvLYvIlePheThrSerIeYLYvValThrAerN	380
Db	1126	GAGTTTAGGAAAGGAGTTTGAACAACGTCCTAAATCTTTTGAATAAATTAAGAAATATA	1185
Qy	381	ArgLeuPhearPProThrLYvGIYvAAnAnGIYLeuValGIYTPRGIuGIYLeuProthr	400
Db	1186	CACCTC-----TCTGAATCCGGAAGATCATTCAGATGTATAGTATACGACA	1233
Qy	401	PheLeuSerAerGIuAerPYvAlaLYvLeuGIuSerTYrCYvPheTYrPheGIuLYvLYv	420
Db	1234	TTTCTTAGTGCATATGACTGCACAAGATGTAAAGTACAGCTGTTTTATTTAAAAAGTCAA	1293
Qy	421	CysPro--AerGIYGIuAerNAlaCYvLYvAAnIleArGIAlaThrCYvTYrLYvArGIY	439
Db	1294	GCACCTCTTGCACAAAGAA-----TGTAAATATCTGAAGCAGACAGATTTATAAGAGGG	1347
Qy	440	LeuAerAlaArGIAlaerNLYvAlaLeuGIuInGIuAerMeArGIYmeLeuHIeGIYSer	459
Db	1348	CTTAAGACACAAGCTTAATGAACATTTGACAAAAAAGATGACGAGCTGTCTATGTGTTCA	1407
Qy	460	AerNLYvSerTPLeuGIuLYvPheGIuInGIuLeuValLYvValCYvGIuLYvLeuLYv	479
Db	1408	GGCAAAAGATGTATTAGAAACTACTAGAAAAAATTAATAGGAAGATGTTCGAAACTTAA	1467
Qy	480	LYvGIuAerNLYvGIYSerPheSerAerNArGIuLeuPheIleLeuCYvAlaGIProAla	499
Db	1468	ACA-----ACAAGCATGAGTGTGTTTGTGATGTATGATGACCACTT	1509
Qy	500	LYvAlaAlaArGIuLeuThrHIeAerLeuArGIvThrIlePheLeuArGIuIn	519
Db	1510	AAACACAGTCAAAATCTTGACAGCTGATATTCMAAGCAAGCAAGCTTTTTCGGAAAAACA	1569
Qy	520	LeuAerGIuLYvArGIaerPheProThrxarPYvAAnCYvLYvGIuLeuGIYArGIYCyv	539
Db	1570	TTGGATCAAAAGCCGAGACTTTTCCAAACACACAAAGTTCMAAGAAATTAGAAAGAAAGT	1629
Qy	540	GIaerPheGIYGIuAerSerLYvGIuIleThrTTPProCYvHIeThrLeuGIuInGI	559
Db	1630	GAACTTTAGGGAAGAGATTCAAAATCAGATTAAAGTGCCATGTCAATACGCTAAACAAACAG	1689
Qy	560	CyAaerNArGIuGIYThrThrGIuIleLeuLYvGIuValLeuLeuAerGIuHIeLYvAer	579
Db	1690	TGTGATCCCTTGGGACATACAGAAATCTTGAACAGGTTTATCTAAGATGAACACAAAGAT	1749

Oy	580	ThrlleuylsaapnglIngluSerCyValLlystrlyleuylsgllybCybaenlybTrpSer	599
Db	1750	ACITTAAGAAGCTCAGAAAACGTGACGAATATTTTAAAGAGAAAATGTCATTAATGGCT	1809
Oy	600	ArgatrgllysaapAPaCPheSerPheValCyValPhegluaenAlaThrCyGluIleu	619
Db	1810	AGAAAGGGGTGTGATCGTTTCTCTTTGTATGTGTTTACCAAAACGCTACGTGAAGTGTG	1865
Oy	620	MetValLysaPVallysaapAPaCPheSerPheValCyValPhegluaenAlaThrCyGluIleu	639
Db	1870	ATACTAGATGATGTGAAAGACAGGTGTGAAGTATTGAAAAAATATGCAAGCGTCAGAT	1929
Oy	640	IleleaglupheleugluabnaenThrAsnlyleThrThrleuGluaygaPnCyPro	659
Db	1930	ATTAAATATCTCTTTAAAAATTAACAATAAAAACAGAACTCAGCAGCAATATTTGTC	1989
Oy	660	SerTrpHisThrTrpCybaenAPaCPheSerProbaPnCyProglYleuThrLys---	678
Db	1990	TCATGGACCCCAATCTCGATAGATTTTATCCCAATGTCTGTATCTTAAGAAAGAA	2049
Oy	679	AsnSerCyThrlybIlelybLysIleargGluProPheTrpLysargLysalaLeuGlu	698
Db	2050	ACTTCTGTCTCAAAATCTTAAAAATATTCGACACCATTCCTACAAAGAGAGCTTTGAA	2109
Oy	699	AspAlaLeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCyGluProAla	718
Db	2110	GATCTCTTAAGTAGAGCTTCAAGGAAATTTAAGTAATAGAAATTAATGGAATCTGCA	2168
Oy	719	LeuLysargTrpCyThrValAlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCy	738
Db	2170	TTTAAAGATATTCACAMATTTGAAAAATGTAAGTATTCATCAATCAACAGTTTA	2229
Oy	739	LysAlaAsnThrLysaapAsnSerGlyLysSerargLysAlaIargLysGluLeuCy	758
Db	2230	AAAGATATATCCGAAGTAATAAATCTAAAAAGACCAATATGAAGTTAGAAAGACCTTGT	2288
Oy	759	GluLysLeuVallybgluValGluGlnGlyLysAlaLeuProthrGluLeuGlyGln	778
Db	2290	CTAATAATTAGTGAAGAGGTGGACAGCAATGTAAATGTTCACAGCAGAAATTTGAGCAT	2349
Oy	779	ProAlaIleAspLeuLysbAPTrpLysThrTrpGluGluLeuLysLysargAlaGlu	798
Db	2350	GAGCAAAAAAGCTTAAAGATTTTGGAAACATTTGAAAACTTAAAAACAGACAG	2409
Oy	799	GluAlaMetAsnLysSerSerLeuValLeuSerLeuLysLysAsnGluSerAsnAla	818
Db	2410	AAAACAATGAATTAATCCAAATCTTGTTTATCATTCGTTAAGAAAGATGAAAAATATATCA	2468
Oy	819	SerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGluAspThr	838
Db	2470	TGCAAAAAATAGTACCAAAAGACAGATAGAAATCCGTTCAAAACGACCTTCAAGATACC	2529
Oy	839	ThrLysIleVallybIleLeuLysargGlyValLysaPValSerValThrGluLeuGlu	858
Db	2530	ACAGAACATATGAATAATATCTACGAGAGAGAGTTAAGCATGTATCCGTAAACGAATCTGAA	2589
Oy	859	AlaLysValaPheAspLeuAlaAlaGluValPheGlyLysargTrpValaAspLeuLysGluLys	878
Db	2590	GCTAAGGCAATTTGATTTGGTAGCAGAAATTTTGAAGATATCTTAAGCTTGAAGAAAGA	2649
Oy	879	CysAsnLysLeuGluSerAPaCPheArgIleLysGluAspCybLysaPLeuGluGluVal	898
Db	2650	TGTATATTAATTTGGAAATCGATTGCGAGCTTAAGAGAGATTTCAAGAGATTTTGAAGAGATA	2709
Oy	899	CysbLysLysIleAsnLysAlaCyargAsnLeuLysProLeuGluValLysProHisGlu	918
Db	2710	TGTGAAAAGATACAGAGAGATGTTCGAAATTTAAAAACAATGAAGAGAACCCCAAGAA	2769
Oy	919	ThrValThrGluSerThrThrThrThrThrThrThrThrThrValAlaAspProLys	938
Db	2770	ACAATGACAGAAACACAAACGACGACACGACGACAAACAGACCGTTATCTGAATCCGAA	2829

```
OY 939 AATHTGTCyGlySerLeuGlnThrThrAspThrTrpValThrGlnThrSerThrHis 958
|||
DB 2830 GCACAGAGATGCAGAACTCTTACAGACCAAGATACATGATGATACACAGACTTCGACACAT 2889
|||
OY 959 ThrSerThrSerThrIleThrSerThrIleThrSerLeuIleThrLeuThrSerThrArg 978
|||
DB 2890 ACCAGCAGCTTACATCATCATCTTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2949
|||
OY 979 ATGCTyGlyProThrIleCysThrThrIleGlyGlnIleuAspAspAlaGlyAspValIlePro 998
|||
DB 2950 CATTGCAACCAACCAACCAAGTGTGACAGAGG-----GATGATGACAGAGAGCTGAAAGCG 3003
|||
OY 999 SerGlnGlyLeuArgMetSerGlyTrpAsnValMetArgGlyValIleValAlaMetVal 1018
|||
DB 3004 AGTGGAGGATTTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3063
|||
OY 1019 ILeSerPheMetIle 1023
|||
DB 3064 ATTTGCTTCATGATTT 3078
|||

RESULT 6
US-09-762-724-1
; Sequence 1, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3042
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3042)
; US-09-762-724-1

Alignment Scores:
Pred. No.: 5,316-290 Length: 3042
Score: 2762.50 Matches: 541
Percent Similarity: 69.10% Conservative: 177
Best Local Similarity: 52.07% Mismatches: 280
Query Match: 51.05% Indels: 41
DB: 4 Gaps: 18

US-10-654-416-14 (1-1023) x US-09-762-724-1 (1-3042)
OY 1 MetAlaArgAlaValIleArgGlnAlaAlaGlyThrGlnAsnSerIleAspGlnGlnHis 20
|||
DB 1 GTGGCGCGGCGGCTTAAGCGGCAAGTAACAGAGACA---TCAGAGATGATGATGAGAGAA 57
|||
OY 21 Val-----LeuAlaLeuIleLeuIleuArgGlyLeuSerGlnGlnGlyCysLeuIle 38
|||
DB 58 GTGCGCTCTTTGGCTTTAATCTTAATAAGAAAGATTCTTAAGATGATTAATAAATGCGAAGAA 117
|||
OY 39 LysLeuLysLysTyrCysGlnGlnIleuThrGlnAlaLysLeuAsnIleGlnGlnValHis 58
|||
DB 118 AATTTAGAAACATTTGCAAGAAATTAAGTGAAGCAATCTACTCCAGAACAGATACAT 177
|||
OY 59 ArgLysLeuLysGlyPheCysGlnAspGlyLysAlaAspThrLysCysLeuGlnIleuLys 78
|||
DB 178 GAAAGATTAAAGATTTCTGTGATGCAAAACGATTAATAAATGTAAGAACTAATAA 237
|||
OY 79 AlaAsnIleGlnLysLeuCysThrThrIleLysGlyLysLeuLysGlnAlaIleLysLys 98
|||
```

```
DB 238 AAAATGTGAAAAAATGCGGTGATTTTAAACAAATAATTGAGAAATTTGCTAATAAG 297
|||
OY 99 LysIleGlnIleIleThrAspLysAspCysLeuGlnAsnGlnGlnIleCysLeuPheLeu 118
|||
DB 298 GAACCTTCAAAATTTGAAAAATGATGAGTGTACAAAAATGACAAACAGCTGCTTTTAA 357
|||
OY 119 GlnGlyValCysSerLysGlyLeuLysAspAspCysAsnThrIleuArgAsnLysCysTyr 138
|||
DB 358 GAAGAACAGCTCTCT---GATCTTACAAAGAAATGCAACGATTTTAACAAATGTTAT 414
|||
OY 139 GlnLysLysArgAspLysValAlaGlnGlnValLeuLeuArgAlaLeuArg----- 155
|||
DB 415 CAGATTAAGCTGTAAAGTGAAGCAAGAGCTTTTAAAGATTAATAAGAAAGAAAT 474
|||
OY 156 ---SerAspLeuAsnGlySerValIleCysGlnLysLysLeuLysGlnIleCysProVal 174
|||
DB 475 TTTAAGATTAATAAATTTCA-----TGCGAAATTAATCGAAGATACGTCAAGAA 525
|||
OY 175 MetGlyArgGlnSerAspGlnLeuThrAsnLeuCysLeuAsnGlnLysGlnThrCysLys 194
|||
DB 526 TTAAGTCAAAATGAGTGAAGATTCATGATGAAATTATGTTTGTCAAAAAAATACGTGAT 585
|||
OY 195 AsnIleLeuIleGlnLysAspLysLysCysGlyThr-----LeuLysThrAspVal 211
|||
DB 586 AATCTTGTAAGAAAGCAACCAAGGTGATCTTTCAAAAATCTTAAACGGAATTT 645
|||
OY 212 SerAlaIleuGlnGlySerPheLys---LysGlnThrCysLeuGlnLeuGlnGlnCys 230
|||
DB 646 AAACCAATTAAGCAAGATGAACACATTAATAAATAAATGCCATTTATTAATGAAGATGC 705
|||
OY 231 TyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGlnLeuGly 250
|||
DB 706 ATTTTATGATGTAAGATGTGGAAACGATTTCA---CTGAAGTGTAGTAATGAAATA 762
|||
OY 251 LysCysGlnGlnGlnIleAlaIleTyrMetProGlyProAspPheAspProThrArg 270
|||
DB 763 AAATGTCAAGAAATAATTAATTAACATTAATCAATTAATCAGGATTTGATCTTAAGAA 822
|||
OY 271 ProGlnAlaThrIleAlaGlnAspIleGlyLeuGlnGlnPheTyrLysValIleGln 290
|||
DB 823 CCAGAAATTAATTAATGCAAGAAAGATGACTTAAGAAATTTATTAAGAAAGCAAGAA 882
|||
OY 291 AspGlyValPheIleGlyLysAsnIleLeuArgAspAlaThrAlaLeuLeuAlaLeu 310
|||
DB 883 GAAGCAACTCTTTGTGGGAAACCTTTACCAAGCAATGCTACTGCTTGGCGCATTTTGG 942
|||
OY 311 IleGlnAspSerSerLeu---LysLysLysAspAspLysGlnLysCysGlnGlnAlaLeu 329
|||
DB 943 ATTCAGATCCATCTTTCAACATCAACGAACTAACAAGAAATGTAATAAATAATTTCT 1002
|||
OY 330 GlnLysSerCysLysAsnProIleGlnIleGlnAlaLeuGlnIleCysLysAsn 349
|||
DB 1003 GAAGATTAATGTAAATTTTAAAGAAACATGATTTTAAAGAAAGCTTAATGCGAGATTAAT 1062
|||
OY 350 GlyLeuSerAsnAspGlyThrLysLysCysGlnLysLeuGlnIleAsnAspIleAsnLysThr 369
|||
DB 1063 AATCCAAATTAAGGATTAAGCAAAAAAATGCCAAGAACTTACTACAGATATTGAAGAAACA 1122
|||
OY 370 CysLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspProThrLysGlyAsn 389
|||
DB 1123 TGTAAATTTTTCATTTCAAAAAACCTTAATGATTCATTTTGTGGCATGCAAAATTAAT 1182
|||
OY 390 AsnGlnIleValIleGlyTrpGlnGlyLeuProThrPheLeuSerAsnGlnLysCysAlaLys 409
|||
DB 1183 GATGGAATTAATTAATGCGGGAATTTATCAACGTTTCTAAGCAATTAAGATTTGACAAA 1242
|||
OY 410 LeuGlnSerTyrCysPheTyrPheGlnLysLysCysProAspGlyGlnAsnAlaCysLys 429
|||
DB 1243 TTAGAAATCGATGTGCTTTATTTTGAATAAAGCTGT---AGAACGAAATCGCATGCAAG 1299
|||
OY 430 AsnIleArgAlaThrCysTyrLysValGlyLysAspAlaAspAlaAsnLysValLeuGln 449
|||
```

Db 1300 AATATCAGAGCAGCATGCTACAGAGGACCTTGACACATTAGCAAAATGATTATTCAA 1359  
 Qy GluabmefaraglymetleuH1eglySeranlyserTrpleuGluLyPheGln 469  
 Db 1360 AAGAAATGCGAGAAATGCTCATGTTCAAAATTAACATGCTTACTAGGTTTCCAAAA 1419  
 Qy 470 GluLeuValyVal1CyegLulyleuLylysgLubnlysgLyserPheSerAasn 489  
 Db 1420 AAATCTATAGAAAGTGTGCAAAAAGTGAAGAAATGAAGATTAAGAGTTTCCAGATAT 1479  
 Qy 490 GluLeuPhe1leuCyeva1GlnPro1Ala1ala1leuLeuThi1AaspLeu 509  
 Db 1480 GAATTAATTTGCTATGTGTCAACCATCAAAAGACGCTCGATTGCTTCCATGATCTT 1539  
 Qy 510 ArgmetLeuThri1ePheLeuH1egLlnleuAaspGlnLyAsgAaspPheProThiAasp 529  
 Db 1540 CCGATGAAGAAATATCTTTTTCAGAAAGACGATTGAAACGAAAGCGAGATTTTCCAGTAAA 1599  
 Qy 530 LyAasnCyelysgLuleuGlyArgLyseCyegLlnAaspLeuGlyAaspSerLyseGlu1le 549  
 Db 1600 GAAGACGCGAGAAATTAATTAAGAAATGTGAGGCTCAAGAAAGATTCATAAAAAATT 1659  
 Qy 550 ThrTrpProCyH1sthrLeuGlnGln1CyAasnArgLeuGlyThrThiGlu1leu 569  
 Db 1660 GAATGGCCATGTCAATCACTTAAGCCAAATTTGTATCAATTAAGAAACGCTAAAGAAATTG 1719  
 Qy 570 LyegLlnValleuLeuAaspGln1lelyAaspThriLeuLyAaspGln1userCyeva1lys 589  
 Db 1720 AAGAAACCTTTAATTAATGAACATTAAGATTAATTAAGAAATTAAGAAATTTGGAATG 1779  
 Qy 590 TytleuLyegLulysCyAasnlyserTrpSerArgArgGlyAaspAaspPheSerPheVal 609  
 Db 1780 TATTGAAGAGAAATGCAATGAATGCTTGAAGAGAAATGAACGTTTCTCTTTTA 1839  
 Qy 610 Cyeva1PheGlnAasn1a1thrCyegLuleuMetValyAaspVal1LyAaspArgCyegLul 629  
 Db 1840 TGTCCTTTGCAAAATAGACCTTGCAAAATTAATGTGTAAGAAATGTGAAACCAATGCAAA 1899  
 Qy 630 ValPheLylyAasn1lelyA1aserTrile1egLlnPheLeuGlnAasnThiAasn 649  
 Db 1900 ATATTGAAAAAAACATTAAAAATACCAAGTATGAT-----AGTAAAACT 1947  
 Qy 650 Lyse1leThrThriLeuGlnAaspCyAaspProSerTrpThiThriThriCyAasnArgPheSer 669  
 Db 1948 ABAATTAAGAAATTAAGGACATATGCTTATTTGGCAACCAACGTCGATCAATGATTTGGA 2007  
 Qy 670 ProAasnCyAaspProGlyLeuThriLyseGlnAasnSerCyThriLy1lelyValyH1eArgGlu 689  
 Db 2008 CCCAATTTGCCCGGATCTTGAAAAA---AATAAATGTGAAGACTTTGAAAAAATATTGCAAA 2064  
 Qy 690 ProPheTrlylyAsglyVal1leuGlnAaspAla1leuLyVal1GluLeuGln1Lylyleu 709  
 Db 2065 CCTTAATTAAGAAAGACCTTGAAAAATGACCTTAATTAATTTGAGTTAGGACATCTT 2124  
 Qy 710 ThrAaspLyserLyCyegLulPro1Ala1leuLyAaspTrpVal1Agl1AasnVal 729  
 Db 2125 GATTAAGAAAAAACTGCAAAACAATCTTGATAAGTCTTACACTTAATGGAATCAAAACA 2184  
 Qy 730 AasnAasn1aser1leSerGlyLeuCyelyVal1AasnThriLyAaspAasnSerGlyLyser 749  
 Db 2185 GGAATTAATAACACTTAAGGTTTGTGAACAGTTCTACTGATTAACAAT----- 2232  
 Qy 750 AaspGlnAaspAla1aglyGluLeuCyegLulysLeuVal1LyseGlnVal1GluGln1Cys 769  
 Db 2233 ---GAAACATTTAAGATTAACCTTTGCAAAAACATGATTCAGCGGTGAAAGAAAAATGC 2289  
 Qy 770 LyAalAleuProThiGlnLeuGly1GlnPro1Ala1AaspLeuLylyAaspTrlyLeuThr 789  
 Db 2290 CAAGATTAATCAAAAGAACTTGAAGAAACAACAAAAATGATTTTGAAGAAAAACATTAAGAT 2349  
 Qy 790 TyrgLulGluLeuLylyAsgAlaGln1Ala1AasnLyserSerLeuVal1leuSer 809  
 Db 2350 TATGAAAAAAGTAAAAAAGATTAACAAAAATGCAATGAAAGAAACAAATCTCGTTTTTCA 2409

Qy 810 Leu1lelylyAasnGln1userAasnVal1SerLyseSerAasnSerLylyAasnLyAaspLyAasn 829  
 Db 2410 ACAACTTAATCAACAGATTAATAAACAAGAAAAAGG-----GTCAGACCT 2454  
 Qy 830 AlaVal1SerAasnGln1leuGlnAaspThriThriLyH1eValy1leuA1aglyGlyVal 849  
 Db 2455 AGTAAGCCTGTGTAGTCAATCAAGATATTGTA---CATTTTAACCTTTGTAAGAAAGATGA 2511  
 Qy 850 LyAaspVal1SerVal1ThriGlnLeuGln1Ala1AaspPheAaspLeuAla1Agl1Val1Phe 869  
 Db 2512 AAA---GTTCAAGTGAAGAAAAAGAAAGCGTTTAATTTGTAAGCATTAAGCATTC 2568  
 Qy 870 GlyArgVal1AaspLeuLyseGlnArgCyAasnLyseGln1userAaspCyAarg1lelys 889  
 Db 2569 AGCTTTATGTATGAGCTTAAGAAAGAAAGCTGTACACATCTAAAGATGATTCGAAATTTAGA 2628  
 Qy 890 GluAaspCyelyAaspLeuGlnGlnVal1CyelylyVal1leuA1aglyVal1AaspAasnLeu 909  
 Db 2629 AAAGAAATGTAAA---TGTAAGAACAGTGCAGAAAGATGAAGAAAAAATATGTTAAAAATA 2685  
 Qy 910 LyAProLeuGln1ValylyAaspProH1sgLulThriVal1PheGln1userThriThriThriThr 929  
 Db 2686 GAACCACTGAAGTAAAGCCACATGAATTAATAAAGCATTAACGGAACCAACATTAACACA 2745  
 Qy 930 ThrThriThriVal1AlaAaspProLyA1a1Thr-----GluCyelylySer 944  
 Db 2746 GTCACAGAAACAGTCAAGAAAGACGAAAGAAACAGTGAAGACGAGAGAAATGCAATCT 2805  
 Qy 945 LeuGlnThriThriAaspThriThriVal1ThriGlnThriSerThriThriSerThriThri 964  
 Db 2806 CTCAGCACAACAGACAGTGGTCAACAAAGCCTCAACAGCAGACCTCCACACT 2865  
 Qy 965 ThriSerThri1eThriSerLyly1eThriLeuThriSerThriArgArgCyelyAaspProThi 984  
 Db 2866 ACCTCCAGCTTACGTCAAGAAATTAACCTGACGAGGCGGTGTAAAGCTTAAGCGAAG 2925  
 Qy 985 CyethThriGlyGlnGlnAaspAaspAla1AaspVal1LyAaspProSerGln1LyLeuArgMet 1004  
 Db 2926 TGTACGACAGAGAGAGAAATGAAGACGAGAGAGTGAAGCGAGTGAAGGCGGTGAGGATG 2985  
 Qy 1005 SerGlyTrpAasnValMetArgGlyVal1leVal1AlaMetVal1leSerPheMetIle 1023  
 Db 2986 AGTGGGTGAGTGTGATGAAGAGGGGTGTATTAGCAATGATGATTCATTATCATGAT 3042

RESULT 7  
 US-09-762-724-3  
 ; Sequence 3, Application US/09762724  
 ; Patent No. 6664053  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kovacs, et al.  
 ; TITLE OF INVENTION: Identification of a region of the major surface  
 ; FILE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii  
 ; FILE REFERENCE: 4239-58054  
 ; CURRENT APPLICATION NUMBER: US/09/762,724  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750  
 ; PRIOR FILING DATE: 1999-08-17  
 ; PRIOR APPLICATION NUMBER: US 60/096,805  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 3006  
 ; TYPE: DNA  
 ; ORGANISM: Pneumocystis carinii sp. f. hominis  
 ; FEATURES:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(3006)  
 US-09-762-724-3  
 Alignment Scores: 5.37e-251 Length: 3006  
 Pred. No.:

Score: 2404.50 Matches: 479  
Percent Similarity: 65.37% Conservative: 195  
Best Local Similarity: 46.46% Mismatches: 320  
Query Match: 44.44% Indels: 37  
DB: 4 Gaps: 19  
US-10-654-416-14 (1-1023) x US-09-762-724-3 (1-3006)

QY 1 MetAlaAGAlaValAlaLysArgGlnAlaAlaGlyThrGlnAsnSerTLeAspGluGlnHis 20  
DB 1 GGGGGCGGGCGGGTCAAGCGCGGGCTGCA---GCACAGAAATGATTGTAAGAAACAAATAT 57  
QY 21 ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGlnCysValLysLysLeu 40  
DB 58 CTTTGGCTTTGATTTTGAATAATGATGCAAAATGATGATTAATGTAATAAAAGGTTG 117  
QY 41 LysLysTyrCysGlnGluLeuThrGluAlaLysLeuAsnIleGluGlnValHisArgLys 60  
DB 118 AAAGAGTATTGTGAAGTTTAAAAATGTAACAAAGAAACCAAAAACCTAGAACAAAAAG 177  
QY 61 LeuLysGlyPheCysGluAspGlyLysAla---AspThrLysCysLysGluLeuLysAla 79  
DB 178 TTAGACGAAATCTGCAAAAGATGATTAACATAGAACAAATGCAAAAGATTCAGAAACA 237  
QY 80 AsnIleGluLysLysCysThrThrLysGlyLysLeuLysGluAlaIleLysLysLys 99  
DB 238 AAGGTTAAAGCAAAATGTAATGTTTTCACACAGAACTTGATTAAGCGAAAGGCA 297  
QY 100 IleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGln 119  
DB 298 GCTTCACATTAAGATTAATGATTTGTAAGAAATGAAACGACATGCTGTTTTTGAG 357  
QY 120 GlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuAspAsnLysCysTyrGln 139  
DB 358 GGAGCATGTCCAAACGAACTTAAGATTAATGATGAACGAGAAATTAATGTTATCAA 417  
QY 140 LysLysAspAspLysValAlaGluGlnValIleLeuArgAlaLeuArgSerAspLeuAsn 159  
DB 418 AAAAAACGAGAGACGTAACGAAACAAAGCTCTTTTAAGAGTACTTAAGAGGAACTTAAG 477  
QY 160 GlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGluSer 179  
DB 478 GATAAAAAACATGCAAAATTAAGTTAAAGGGGGTTTGCAGAAATTCACAAAGAAAGT 537  
QY 180 AspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGlu 199  
DB 538 GATGAGCTAAATTAATTAATGCTTGAACGAAAGAAACGCTGAGATCTGTATCTAAG 597  
QY 200 LysAspLysLysCysGlyThrLeuLysThrAspValSerAlaIleLeuGly-----Ser 217  
DB 558 AAAGAAATCAAAATGCAAACTTCAAGAGAAATGATCTAGTCTTGAAAGGAAAGAT 657  
QY 218 PheLysLysGluThrCysLeuGlnLeuLeuGlnGlnCysTyrPheTyrIleGlyAsnCys 237  
DB 658 TTATTTAAAGAAATAATGTTTATTTATTTCTGAAGATGTTTCTTTTAAAGGCTCAACGT 717  
QY 238 GlyAspAspAspIleIleLysCysIleGluLeuGlyLysCysGlnGlnGlnAsnIle 257  
DB 718 ---GAAACAGATCAACCAAGTGAAGGTTTGCAGCAAAATGCAAAAGGAAATCTC 774  
QY 258 AlaTyrMetProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGlu 277  
DB 775 GTTTATGACGACCGAGTTTCACTTGTCTTACGAAATTAAGATTAAGTTAGTTACAGAA 834  
QY 278 AspIleGlyLeuGlnGluPheTyrLysLysValGluGluAspGlyValPheIleGlyLys 297  
DB 835 GAAATAGACCTTAAGAAATTAATGTAAGACGAAAGGAAATTCATCATTTGGAAG 894  
QY 298 AsnHisLeuAspAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLys 317  
DB 895 CCATCAATTAAGAGTCAAGTCCCTTATTTGCAATTAAGCAAGAGTGAATGCT----- 948  
QY 318 LysLysAspAspLysGluLysCysGluGlnAlaLeuGlnLysSerCysLysAsnProHis 337

DB 949 CAAATACTTTTAAAGATCAATGTAAGATGTTTAAAAAAAATGCAAACTTTAA 1008  
QY 338 GluHisGluAlaLeuGluSerLeuCysValLysAsnGlyLeuSerAsnAspGlyThrLys 357  
DB 1009 GAGCATATTATTAAAAAGATTTATGTATGTAATGAAGACTTACCTGATTAATCAAAAGAA 1068  
QY 358 LysCysGluGluLeuGlnAsnAspIleAsnLysThrCysLysIlePheThrSerLysVal 377  
DB 1069 AAATGCGAAGAACTTAATTAAGAGTTA-----ACAACCCGTATTTTAACCTTTTCAA 1122  
QY 378 ThrAsnAspArgLysPheAspProThrLysGlyLysAsnGlnIleValGlyTyrGluGly 397  
DB 1123 AGGATTGAGAAATATTTCGCTCCAGCT---AATGTAAGGAAATTAATGCTGGCATATG 1179  
QY 398 LeuProThrPheLeuSerAsnGlnAspCysAlaLysLeuGluSerTyrCysPheTyrPhe 417  
DB 1180 TTGCATACATTTCTTGGTGAAGAGAGTGAAGCAAACTGTTGCGATGTTGTTTATTTG 1239  
QY 418 GluLysLysCysProAspGlyGluAsnAlaCysLysAsnIleAlaGlyThrCysTyrLys 437  
DB 1240 AAAAGCCAAAGCTCA---CTTGAAAGCCCTGCAATTAATTAAGACAGCATGTTATAA 1296  
QY 438 ArgGlyLeuAspAlaArgAlaAsnLysValLeuGlnGlnGluAsnMetArgGlyMetLeuHis 457  
DB 1297 AAAGGCTTGAAGACGTAGCAAAAGCAATTAAGACATTAAGTAAACGGGAAATTCGAA 1356  
QY 458 GlySerAsnLysSerTyrLeuGlnLysAspPheGlnGlnGlnGlnLysValLysGluLys 477  
DB 1357 GGTCAATAGAAACATGCTGAAGAACCTTCAAAAAAAGCTGTAAGAAAGTTTGAAAG 1416  
QY 478 LeuLysLysGluAsnLysGlySerPheSerAspAspGluLeuPheIleLeuCysValGln 497  
DB 1417 ACGAAGAGGAA-----AGTATGAATTTATTTGATCTATGTAATGATCAAC 1458  
QY 498 ProAlaLysAlaAlaArgLeuLeuThrHisAspLeuArgMetLysTyrIlePheLeuArg 517  
DB 1459 CCAATTAAGAAAGGCTCTTACAGTGTCAACAGATTTGGAGATGAGGCAAGTTGCTTGA 1518  
QY 518 GlnGlnLeuAspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArg 537  
DB 1519 GAGCATTTGAAGCAAAAGCAAGATTTTCAACAGAAAGATGTTAAAGAAATTAAGAGAA 1578  
QY 538 LysCysGlnAspLysGluLysAspSerLysGluIleThrTyrProCysHisThrLeuGlu 557  
DB 1579 AAATGTAAGCTTGAAGAAAGATTAAGAAATTAATGATGATGATGATTAATGATTA 1638  
QY 558 GlnGlnCysAsnArgLeuGlyThrThrGluIleLeuLysGlnValIleLeuAspGluHis 577  
DB 1639 CAGCATTCGAAATCGGCTGAAGACATGAGCAGCTTAAGAGAGAGTTGCTTAAGAAAT 1698  
QY 578 LysAspThrLeuLysAspGlnGluSerCysValLysTyrLeuLysGluLysCysAsnLys 597  
DB 1699 AAAGATATTTAAGATGAATAATGCTGCAAGAGAAAGCTTAAGAAACGATGTAAGAA 1758  
QY 598 TyrSerArgArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCys 617  
DB 1759 TGGTTTGAAGAAAGAAATTAATTAATTTTTCGCTGTGCTGACTTGAACTTGTTGG 1818  
QY 618 GluLeuMetValLysAspValLysAspArgCysGluValPheLysLysAsnIleLysAla 637  
DB 1819 AAAAAAGTCACTAGAAATGTTGAATCTTAATGTAATTAATTAAGAAAGACATATGAAACT 1878  
QY 638 SerTyrIleIleGluPheLeuGluAsnAspThrAsnLysIleThrThrLeuGluArgAsn 657  
DB 1879 ATGAAGCTTAATTAAGTAAATGCTTAAGAAAGAGAAATA----- 1920  
QY 658 CysProSerTyrPheThrTyrCysAsnArgPheSerProAsnCysProGlyLeuThr--- 676  
DB 1921 TGTGAATTTTGGCTCATATTGTTAAAGATGAAGCAAAATTTGTGAAGAACTTAAGAAAC 1980  
QY 677 -----LysGluAsnSerCysThrLysIleLysLysHisArgGluProPheTyrLysArg 694

```

Db      1981 GGAGGAAAAGATGGCAATGCAAAAACTCAATATAAAGTCMAATCTTCCTTGA AAAA 2040
Qy      695 LysAlaLeuGluAaPAlaLeuValGluLeuGlnGlyLysLeuThrAaPLeuSerLys 714
Db      2041 GAACTTTAGAAAATAAAGATTGTGAAAGAAATTTGAAAGAGTACTTATATCAAAACGTAGAGAA 2100
Qy      715 CysGluProAlaLeuLysArgTyrCysThrValAlaGlyAaPValAaPAlaSerLys 734
Db      2101 TGTAAACATACACTTAATATATATACGTACAAATGAAAAAGCAGAGAAAT--GGGTG 2157
Qy      735 SerGlyLeuCybLysValAaPThrLysAaPLeuSerGlyLysSerAaPLeuAaPAlaArg 754
Db      2158 GAAACTTTGTGCAAA-----AGCAAAAGAAAACACC--AAGAGTGCATTAAAGTTAGA 2208
Qy      755 LysGluLeuCybGlyLysLeuValGluValGluGlnGlnGlyLysValAaPLeuProThr 774
Db      2209 GAAAGACTCTGTGAAAGCTTATATAACGTATATAAAGAAATATCTCAAAATTAAGAC 2268
Qy      775 GluLeuGlyGlnProAlaAaPLeuLysLysAaPValAaPLeuThrTyrGluGlnLys 794
Db      2269 GAGCTTGAGAGAAAGTAAAGAGCTTTAGAAAAGAAAAGAAAGTATAAATAAATAA 2328
Qy      795 LysAaPAlaGluGlnAlaMetAaPLeuSerSerLeuValLeuSerLeuLysLysAaP 814
Db      2329 GAAAGAACAGAAAAGCCATGGAAGATGCAAACTTATTTATCGAAGCGAAAGACT 2388
Qy      815 GluSerAaPValSerLysSerAaPLeuLysAaPValAaPAlaValSerAaPVal 834
Db      2389 GATATATATATAATTAATAG-----TCAGTAAATATAAGCTCATCTGATACCTAAGAA 2442
Qy      835 LeuGlnAaPThrThrLysValValValLysValLysValArgGlyValLysAaPValSerVal 854
Db      2443 GGAAGAAAGCACAACA-----GATTTAACTTGTAGAAAGAAATGCAAAA--GTGCAGTGA 2496
Qy      855 ThrGluLeuGlnAlaLysAaPLeuAaPAlaGluValPheGlyAaPValAaPVal 874
Db      2497 ACAGAAAAGAAATATAGACGATTTGATTTGTTGACAGAGCATTTGATCTCTACTAGAA 2556
Qy      875 LeuLysGluAaPValAaPLeuLysSerAaPValAaPValLysGluAaPValAaPVal 894
Db      2557 TTGAAAGAAATATATATATCATTTCACTGAGAAATGTGGTTTCAAAAAGAGTGT--GAC 2613
Qy      895 LeuGluGlnAlaLysLysLysValAaPLeuLysAaPValAaPLeuLysProLeuGluVal 914
Db      2614 TGTGAGAGATCATTTAAAGATACAGGAAATATGTTCAACATTAGAGCCACTAAAGTG 2673
Qy      915 LysProHISGluThrValThrGlu-----SerThrThrThrThrThrThrThrThrThr 932
Db      2674 AGACCAACACGAAATATAGTAAATAAACAATCACTACACCAACACACACACCACT 2733
Qy      933 ThrValAlaAaPValAaPLeuThrGluCybLysSerLeuGlnThrThrAaPValThrVal 952
Db      2734 ACCATTAAAGCGAAAGGCAACAGACTGCACCTTTACGAAACAGATACGTGGGTC 2793
Qy      953 ThrGlnThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 972
Db      2794 ACAAAAGCGTGCACCATCTACGACATCCACAACACATTTACAGTCACGTCAAGATA 2853
Qy      973 ThrLeuThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 992
Db      2854 ACCTTGACCTGCAACAGCGGTGTAAGCTTCAAGATGTAAGACAGAGAGAGAGATGAA 2913
Qy      993 AlaGlyAaPValLysProSerGluGlyLeuAaPLeuSerGlyTTPAaPValAaPVal 1012
Db      2914 GCAGAGAACGTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2973
Qy      1013 ValIleValAlaMetValIleSerPheMetIle 1023
Db      2974 GTGTATATAGCAATGACGATTTCAATTCATGAT 3006

```

RESULT 8  
US-07-781-034-1  
Sequence 1, Application US/07781034

```

: Patent No. 5442050
: GENERAL INFORMATION:
: APPLICANT: Fishman, Jay A.
: TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
: TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/781,034
: FILING DATE: 19911018
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MGH91-02A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 861-6240
: TELEFAX: (617) 861-9540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2814 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1617..2813
: US-07-781-034-1
:
: Alignment Scores:
: Pred. No.: 1,666-88 Length: 2814
: Score: 913.00 Matches: 211
: Percent Similarity: 53.99% Conservative: 107
: Best Local Similarity: 35.82% Mismatches: 223
: Query Match: 16.87% Indels: 48
: DB: 1 Gaps: 13
:
: US-10-654-416-14 (1-1023) x US-07-781-034-1 (1-2814)
:
: Qy      5 ValLysArgGlnAla-----AlaGlyThrGlnAaPLeuSerLysAaPLeuVal 21
: Db      1131 GTTAAGAGCGCAACAGACAAAGTATCAAGATGATGATGATTAAGAGGAACTT 1190
:
: Qy      22 LeuAlaLeuLysLysGluAaPValLysSerGluGlnGlyLysLysLysLysLys 41
: Db      1191 TTGGCTTCACTTGTGAAAGCAAAACATGATGATGATGATGATGATGATGATGATGAT 1250
:
: Qy      42 LysTyrCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 61
: Db      1251 GAATATTTGTAAGAGTTGAAAGAAAGCAAGATGAGATTTCC--AGTGTGAATGAGAAAGTT 1307
:
: Qy      62 LysGlyPheCysGluAaPValLysAaPValLysCysLysGluLeuLysValAaPVal 81
: Db      1308 AAAGAGACTTGTATGATGATTAATAAAGCAAGCAAAATGCAAAAGCAAAAGCAAAAGTT 1367
:
: Qy      82 GluLysLysCysThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 101
: Db      1368 AAAGATGAATTTGGAAGCTTTGATACGATCTTGAAGCATCGGTAGAT-----GAC 1418

```





```

: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1617..2813.
: PCT-US92-08328-1

Alignment Scores:
Pred. No.: 1,666-88      Length: 2814
Score: 913.00          Matches: 211
Percent Similarity: 53.99%      Conservative: 107
Best Local Similarity: 35.82%    Mismatches: 223
Query Match: 16.87%            Indels: 48
                                   Gaps: 13

US-10-654-416-14 (1-1023) x PCT-US92-08328-1 (1-2814)

Qy      5  VallyaArgGlnAla-----AlaGlyThrGlnAasnSerIleAapGluGlnHisVal 21
Db      1131 GTTAAAGAGCAAGCAAGTAGTCAAGTAGCACAAAGATGATTAAGAGCAACCTT 1190

Qy      22  LeuAlaLeuIleLeuLeuGlyLeuAapGlyLeuSerGluGlnGluCyblyblyblyblyblybly 41
Db      1191 TTGGCTTTCATGTGAAGCAAAACATGATGATGAAATGATCAAAAAAGGCTCGAG 1250

Qy      42  LyetYrCyegIngluLeuThrGluAlaLyleuAenIleGluGlnValHisArgLysLeu 61
Db      1251 GAATATTGTAAAGCTTGAAGAAAGCAGATGAGATTTC---AGTGTGAATGAGAAAGTT 1307

Qy      62  LysGlyPheCyegLysAapGlyLysAlaAapThrLysCyblyblyblyblyblyblyblyblybly 81
Db      1308 AAAGCACTTTGTGTGATTAATAAAACGACGCAAAATGCAAAAGCACTGAAAAAAAGTT 1367

Qy      82  GluLysblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 101
Db      1368 AAAGCAATTAATGGGAACTTTTGATACGAGATCTTAAGACATCGTGAT-----GAC 1418

Qy      102 IleIleThrAapLysAapCyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 121
Db      1419 ATGCAATGATGAAGAGCTTGTAAATAAAACATGAAGAAATGTATATACCTTTAGAGCAACA 1478

Qy      122 CysSerLysGluLeuLysAapAapCyblyblyblyblyblyblyblyblyblyblyblyblybly 141
Db      1479 GACCAATATAGTCTTAAGAGAACTGTGTCAAGTTGAGGAGGAAGATGTTACGAATTTGAAG 1538

Qy      142 ArgAapLysValAlaGluGlnValLeuLeuArgAlaLeuArgSerAapLysAapGlySer 161
Db      1539 CGTAAAAAGGTGGAGAGAGACTCTTTTGGGCGCTCGAGGGGAGATCTAAAGATGAA 1598

Qy      162 ValIleCyegLysLysLysLeuLysGluIleCyblyblyblyblyblyblyblyblyblyblybly 181
Db      1599 GCTAAATGTAAGAAAGATGAAAACTGTTGCCCAATGTTAAAGCCGAGAAAGTGAAGAG 1658

Qy      182 LeuThrAasnLeuCyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 201
Db      1659 CTGATGTTTTTCTGCTTGAATTCGATGCAAGCTGTAAAGCGCTGAAACAAATACAGAA 1718

Qy      202 LysLysGlyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 221
Db      1719 GAAGTTTGCTGCTTGAAGAAAGAAAGCTTAA-----GATGGCGAATTTAAAGAA 1769

Qy      222 ThrCyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 241
Db      1770 AATGTCATGAAGACCTTGAGAAATGTCATTTTAAACAAAGAAAGGTGATCAAGAA----- 1823

Qy      242 IleIleLysCyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 261
Db      1824 ---ACAAAGTGTAGAGATATAGACATGCAAGAAAGAAAGATTCACATATAAAGCG 1880

Qy      262 ProGluProAapPheAapProThrArgProGluAlaThrIleAlaGluAapGlyIleu 281
Db      1881 CCGGAATCTGATTTTAACTCTGTCAAGCGCTGCTGTTGAGAAAGTATGGGCTTG 1940

Qy      282 GluGluPheTyLysLysValGluGluAapGlyValPheIleGlyLysAasnHisLeu--- 300

```

```

Db      1941 GATGATGTGTATATAAAGCGCTGAAAAAGAAAGAAATTATTGCAAAATACGAGTGGAT 2000
Qy      301 -----ArgAapIleThrAlaLeuLeuAlaLeuLeuIleGlnAapSerLeuLys 317
Db      2001 CTACCAAGAAAGTACAGTACAAAATTTCTCAAGATCTTGTACTGTGAGCAGAGAT 2060
Qy      318 LysLysAapAapLysGlyLysCyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 337
Db      2061 GAGATGATCAGAGAAAGAAATGCGTAAAGCGTTAGAAATGTGAATCTTAAGTAT 2120
Qy      338 GluHisGluAlaLeuGluSerLeuCyblyblyblyblyblyblyblyblyblyblyblyblybly 357
Db      2121 TTGAATACGTATTTGATGAGAGTATACAAA-----GATGCTGATTAAGAAAT 2168

Qy      358 LysCyegLuuGluLeuGlnAasnAapIleAasnLysThrCyblyblyblyblyblyblyblyblybly 377
Db      2169 AATGCAAAAA---AAGCTAGATGTAAGAAAGAGT----- 2204

Qy      378 ThrAasnAapLysPheAapProThrLysGly----- 388
Db      2205 ACAAAACTCAAGTTAAATCTTTATGTGAAGAGGTGTCTACGAGTTTAAAGAGATAA 2264

Qy      389 AasnAasnLysValGlyTrpGluGlyLeuProThrPheLysSerAasnGluAapCybly 408
Db      2265 AATACATCATCTTTATGCTGGGACAGCTTCAACATTAATTACGAAGGAGAGTGTCA 2324

Qy      409 LysLeuGluSerLysCyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 426
Db      2325 GAACCTTGAAGTCGAATTTCTATTAGAAATCGGTAAAGATTAATGAGATTGTGAA 2384

Qy      427 AlaCyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 446
Db      2385 GCGGTCAAAATCTACATACAGCTGCTTATTAAGAGCAACAGCAGAGTCTTAATAG 2444

Qy      447 ValLeuGlnLysMetArgLysMetLeu---HisGlySerAasnLysSerTrpLeuGlu 465
Db      2445 TTCTTCAAAAGAAATGGAAGGAAAGCTTGTCATGATTAAGATTTTAGCGATCTTAA 2504

Qy      466 LysPheGlnGluLeuValLysValCyblyblyblyblyblyblyblyblyblyblyblyblybly 485
Db      2505 GATTGTAAAAAATATGTGTAGAAAACTGTACAAACTTTTAAAAAGATTAAGA----- 2558

Qy      486 PheSerAasnAapGluLeuPheIleLeuCyblyblyblyblyblyblyblyblyblyblyblybly 505
Db      2559 -----TACCTTCAAAATGCTCTTAATCTTAAGAACTATGATGCGCTT 2603

Qy      506 ThrHisAapLysMetLysThrIlePheLeuArgGlnLeuAapGlnLysAap 525
Db      2604 TCAATGATATTTTTCATCAATCAAGAGTTAAGTTCGCTTTAGATGATCAAGAGAT 2663

Qy      526 PheProThrAapLysAasnCyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 545
Db      2664 TTTCCATTTGAAAAAGATTTGCTTGAATTTGGAGAGAGTGTATCAACTTAAGTATGAT 2723

Qy      546 SerLysGluIleThrTrpProCyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 565
Db      2724 TCATTATTTGAATTTAGAAAAGTATTAACATTTGAAAAGAGCGCTGAAATTTTGACGTT 2783

Qy      566 ThrGluLeuLysGlnValLeuLeu 574
Db      2784 ACAGAAAGATTTAGAAAGTATTTTAA 2810

RESULT 10
US-07-781-034-3
: Sequence 3, Application US/07781034
: Patent No. 5442050
: GENERAL INFORMATION:
: APPLICANT: Fishman, Jay A.
: TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
: TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:

```

```

ADDRESS:  Hamilton, Brook, Smith & Reynolds, P.C.
STREET:  Two Millitia Drive
CITY:  Lexington
STATE:  Massachusetts
COUNTRY:  USA
ZIP:  02173

COMPUTER READABLE FORM:
MEDIUM TYPE:  Floppy disk
COMPUTER:  IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/07/781,034
FILING DATE:  19911018
CLASSIFICATION:  424
ATTORNEY/AGENT INFORMATION:
NAME:  Granahan, Patricia
REGISTRATION NUMBER:  32,227
TELECOMMUNICATION INFORMATION:
TELEPHONE:  (617) 861-6240
TELEFAX:  (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH:  1448 base pairs
TYPE:  NUCLEIC ACID
STRANDEDNESS:  double
TOPOLOGY:  linear
MOLECULE TYPE:  DNA (genomic)
HYPOTHEICAL:  NO
ANTI-SENSE:  NO
US-07-781-034-3

Alignment Scores:
Pred. No.:  1,986-50      Length:  1448
Score:  559.00      Matches:  136
Percent Similarity:  55.87%      Conservative:  59
Best Local Similarity:  38.97%      Mismatches:  117
Query Match:  10.33%      Indels:  38
DB:  1      Gaps:  8

US-10-654-416-14 (1-1023) * US-07-781-034-3 (1-1448)
QY 1 MetAlaTgAlaValAlaValArgGlnAlaAlaGlyThrGlnAenSerIle-Aap----- 17
DB 388 ATGGACCGCGCGTTAAGAGCAAGCAAGTAAAGTAAAGACAAAGATGAGATTAA 447
QY 18 -GlUcHisValleuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCyValy 37
DB 448 GGAGGAAACCTTTGGCTTTCATTGTGAAGCAAAACATGATGATGAGAAATGAAATGCAA 507
QY 37 AlAylAspLeuLysLysCysGlnGlnLeuThrGlnAlaLysLeuAenIleGluGlnVa 57
DB 508 AAAAAGCTTCAGAGAAATTTGTAAAGATTGAAGAAACAGATGAGATTTC---AGCT 564
QY 57 lHieArgLysLeuLysGlyPheCysGluAspGlyLysAlaAapThrLysCysLysGluLe 77
DB 565 GAATGAGAAAGTTAAAGACTTTGTGATGATTAATAAAAGAGCGAAATATGCAAAAGACT 624
QY 77 ulAylAlaAenIleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLys 97
DB 625 GAAAAAAAAGCTGGGATGGAATTTGGAACTTTGATACGATCTTGAAGCATCGGTGA 684
QY 97 AlAylAspIleGlnIleIleThrAapLysAapCysLysGluAenGlnGlnGlnCyLeuPh 117
DB 685 T-----GACATAGAAGATGAAGAGCTTTGTAAAAAACCTGAAGAAAAATGATATCT 735
QY 117 eLeuGlnGluValCySerLysGluLeuLysAapAapCysAenThrIleuAgaAenLysCy 137
DB 736 TTTAAGAGAGAGAGAGCAAGCAATAGCTTAAAGAGAGAGCTGTCAAGATTGAGGAGAGATG 795
QY 137 eTyGlnLysLysAaAapLysValAlaGlnGluValLeuLeuAgaAlaLeuAaAaAaAa 157

```

```

DB 796 TTACGAATTGAAGCGTAAAAAGGTGCGAGAGAGCTCTTTTANGGCGCTCGAAGAAAGA 855
QY 157 pLeuAenGlySerValIleCysGlnLysLysLeuLysGluIleCysAProValMetGlyAr 177
DB 856 AGCTAAAGAGAGAGATTAAATGTAAGACAGAGATGAAGAAAGGTTTGCCAGAGTTAAAGCG 915
QY 177 gGluSerAapGluLeuThrAenLeuCySLeuAenGlnLysGluThrCysLysAenIleLe 197
DB 916 AGAAAGGAGAGAAATGATGTTTGTGCTTGATTCGATGCGAAGCGTGCAGCCGTGAAA 975
QY 197 ulleGluLysAapLys-----LysCysGlyThrLeuLysTh 209
DB 976 AAAAAATCAGAAAGAGTTTGCCAGCTTTTAAAGAAACCTTAAAGATGAGCAATTA----- 1030
QY 209 rAapValSerAlaAlaLeuGlySerPheLysLysGlyThrCysLeuGlnLeuLeuGlu 229
DB 1031 -----AAGAAAAATGTCATGAAAGACTTGAGAA 1059
QY 229 nCySerPheTyrlleGlyAenCyGlyAapAapAapIleIleLysCysIleGluLeuG 249
DB 1060 ATGTCATTTTACCGAGAGAGCTGT-----GATAAAAAATAATGTGATGAGATTA 1110
QY 249 yGlyLysCysGlnGlnGlnAenIleAlaTyMetProProGlyProAapPheAapProTh 269
DB 1111 GGATCAATTGCGAGAAAAAAGAAATACATATAA-GCCCGCAATCTGATTTCTAGTCTGT 1169
QY 269 rArgProGluAlaThrIleAlaGluAapIleGlyLeuGlnGluThCyLysLysValG 289
DB 1170 CAAGCCGAGAGAGCGTGTGTTGAGAGATGATGGTGGATGATGATGATGATGATGATGATG 1229
QY 289 uGluAapGlyValPheIleGlyLysAenHisLeu-----ArgAapAlaThrAl 305
DB 1230 AAAAGAGGAATTTATTTGAAAAATCAGAGGTGATCTACCAAGAGAGTCAGATTACAA 1289
QY 305 AlAylLeu-----AlAylLeuIleGlnAapSerSerLeuLysLysAapAapLysG 323
DB 1290 ATTCTGCAAGATCTTCTGCTAGCTTTCAGCAGAGATGAGAAATGATTAAGATGCAAGGAA 1349
QY 323 ulAysCysGlnGlnAlaLeuGlnLys 331
DB 1350 GAAATGCGAAAAAGCGTTAAAAAAA 1374

RESULT 11
PCT-US92-08328-3
Sequence 3, Application PC/TUS9208328
GENERAL INFORMATION:
APPLICANT: Fishman, Jay A.
TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08328
FILING DATE: 19920930
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/781,034
FILING DATE: 18-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,166
FILING DATE: 30-SEP-1991
ATTORNEY/AGENT INFORMATION:

```

```

NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MGH91-02NA PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
TELEX: 951794
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US92-08328-3

Alignment Scores:
Pred. No.: 1.98e-50 Length: 1448
Score: 559.00 Matches: 136
Percent Similarity: 55.87% Conservative: 59
Best Local Similarity: 38.97% Mismatches: 117
Query Match: 10.33% Indels: 38
Gaps: 8

US-10-654-416-14 (1-1023) x PCT-US92-08328-3 (1-1448)

QY 1 MetAlaArgAlaValLlybArgGlnAlaIleGlyThrGlnAsnSerIle-App----- 17
Db 388 ATGCGACGCGCGGTTTAAAGCGCAAGCAACAGTACACAGTAAACACAAAGATGAGATTAA 447
QY 18 -GluglHleValIleuAlaIleuIleleuValAspGlyLeuSerGluGlnGluCysIle 37
Db 448 GGAAGAACACCTTTGGCTTTCATTGTGAAGCAACATCATGATGAGATGCAATGCA 507
QY 37 bLybLybLeuLybLybTyrcyGlnGluLeuThrGlnAlaLybLeuAsnIleGluGlnVa 57
Db 508 AAAAAGCTCGAGGAATATTCTTAAGAGTTGAAGAAAGCAATAGAAATTCT--AGGT 564
QY 57 lHsArxLybLeuLybGlyPheCybGluAspGlyLybAlaAspThrLybCybLybGluLe 77
Db 565 GAATGAAAGAGTTAAAGAGCTTGTGATGATAAACAGAGACGAAAAATGCCAAAGAACT 624
QY 77 uLybAlaAsnIleGluLybLybCybThrTrIleGlyLybLeuLybGluAlaIleLy 97
Db 625 GAAAAAAAAGTTGGGATGGAATTGGGAACCTTTGATACGGATCTTGAAGCATCGTAGA 684
QY 97 bLybLybIleGlnIleIleThrAspLybAspCybLybGluAsnGluGlnGlnCybLeuPh 117
Db 685 T-----GACATAGAGATGAAGAGCTTTGTAAGAAAAACATGAAAAAATGTATACT 735
QY 117 eLeuGluGlyValCybSerLybGluLeuLybAspAspCybAsnThrIleuArgAsnLybCy 137
Db 736 TTTAGAGAGACAGACACCAATACTTAAAGAGAACTGTCAAGTTGAGAGGAAGATG 795
QY 137 eTyrcGlnLybLybArgAspLybValAlaGluGluValIleuLeuArgAlaLeuAspSerAs 157
Db 796 TTACGAATGTAAGCGGTAAAGAGGTGGCAGAGAGACCTCTTTTGAGGGGCGCTGGAAAGGA 855
QY 157 pLeuAsnGlySerValIleCybGluLybLeuLybGluIleCybProValMetGlyAr 177
Db 856 AGCTAAAGAAAGAAATTAATGTTAAAGCAGAGATGAAAAAGTTGGCCAGCTTAAAGCCG 915
QY 177 gGluSerAspGluLeuThrAsnLeuCybLeuAsnGlnLybGluThrCybLybAsnIleLe 197
Db 916 AGAAAGCGACAGAAATGATGTTTGTGCTTGAATCCGATGGAACGATCGCTGAAA 975
QY 197 uIleGluLybAspLyb-----LybCybGlyThrIleuLybTh 209
Db 976 AAAAAATCAGAAAGATTGGCACCTTTTAAAGAAAGCTTAAAGATGCGGAATTA---- 1030
QY 209 rAspValSerAlaIleuGlySerPheLybGluThrCybLeuGluLeuGluGlu 229

```

```

Db 1031 -----AAGCAAAAATGTCATGAAAGACTTGAGAA 1059
QY 229 nCybTyrcPheTyrcIleGlybAsnCybGlybAspAspIleIleLybCybIleGluLeuG 249
Db 1060 ATGTCATTTTAAACGAGAACGGCTT-----GATTAACAAATGATGAGATGA 1110
QY 249 yGlyLybCybGlnGluGlnIleAsnIleAlaTyrcMetProProGlyProAspPheAspProTh 269
Db 1111 GATCAATCGAGAAAAAGAAATACATATTA--GCCCGCAATCTGATCTACTCTCT 1169
QY 269 rArgProGluAlaThrIleAlaGluAspIleGlyLybGluGluGluPheTyrcLybValG 289
Db 1170 CAAGCCGAGACGCTGTTGTTGAGAAAGTATGGGTTGATGATGTGTATAAAGAGCTGA 1229
QY 289 uGluAspGlyValPheIleGlyLybAsnIleLeu-----ArgAspAlaThrAl 305
Db 1230 AAAAGAAAGAAATTTATTTGAAAAATCAGAGTGAATCTACCAAGAAAGTCAGGTACAA 1289
QY 305 aLeuLeu-----AlaLeuLeuIleGlnAspSerLeuLybLybAspAspLybG 323
Db 1290 ATTTCTCAAGATCTCTTCTGATCTTGACGACAGATGAGAAATGATTAAGATGACAGGAA 1349
QY 323 uLybCybGluGluAlaLeuGlnLyb 331
Db 1350 GAATGCGAAAAAGCGTTAAAAAAA 1374

RESULT 12
US-09-762-724-15
: Sequence 15, Application US/09762724
: Patent No. 6664053
: GENERAL INFORMATION:
: APPLICANT: Kovacs, et al.
: TITLE OF INVENTION: Identification of a region of the major surface
: FILE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
: FILE REFERENCE: 4239-58054
: CURRENT APPLICATION NUMBER: US/09/762,724
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: PCT/US99/18750
: PRIOR FILING DATE: 1999-08-17
: PRIOR APPLICATION NUMBER: US 60/096,805
: PRIOR FILING DATE: 1998-08-17
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 15
: LENGTH: 249
: TYPE: DNA
: ORGANISM: Pneumocystis carinii sp. f. hominis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(249)
US-09-762-724-15

Alignment Scores:
Pred. No.: 5.38e-34 Length: 249
Score: 396.00 Matches: 73
Percent Similarity: 98.80% Conservative: 9
Best Local Similarity: 87.95% Mismatches: 1
Query Match: 7.32% Indels: 0
Gaps: 0

US-10-654-416-14 (1-1023) x US-09-762-724-15 (1-249)

QY 941 GluCyLybSerLeuGlnThrThrAspThrTryValThrGlnThrSerThrHisThrSer 960
Db 1 GAGGCGCAATCTCTGACAGACAGACACGTTGGTCAAAAGAGTGCAGACCCATATAC 60
QY 961 ThrSerThrIleThrSerThrIleThrSerValIleThrLeuThrSerThrArgAspCys 980
Db 61 ACTTCTACGACTACGTCACAGTCACATCGAAATTAACCTCACCTTAACGAGCGGTGT 120
QY 981 LybProThrLybCybThrThrGlyGluGluAspAspAlaGlybAspValLybProSerGlu 1000

```



```

Db      3946 GCCCAGAAAGCAGGTTTACACACAGATCAAGTGTTCACAGACAGATGAAAAA 4005
      461 -----LysSerTrpLeuGluLysPheGlnGlnLeuValCysGluLys 477
      4006 GATGCTATGTTTCTCTGCTATCCCGAGCAACCATTTACCAACAGATTGAAAGAA 4065
      478 LeuLysLeuGluLysLysSerPheSerAsnArgLysPheLeuValGln 497
      4066 TTAAAGAGGCGCTAGAA-----GAG 4086
      498 ProAlaLysAlaLysLeuThrHisAspLeu-----ArgMetLysThrIle 514
      4087 GAGACTAAGGCCAAGACACTGTGCCCATGCCCTGACGTACGCCCATGACTGTGAC 4146
      515 PheLeuArgGlnGlnLeuAspGlnLysArgAsp----- 525
      4147 CTGCTGCGGGAACAGTATGAGAGAGACAGAACCCAGGCTGACGTGAGAGGGAATG 4206
      526 -----PheProThrAsp----- 529
      4207 TCCAAGGCCAACAGTGAAGTTGCCAGTGAAGCAACAGTACGAGAGCGCATTCAG 4266
      530 -----LysAsnCysLysGluLeuGlyLysCysGlnAspLeuGly 543
      4267 CGCACAGAGGAGCTGAGAGAGGCCAAGAAAGACTACCCAGCCTCTGACAGATGACAA 4326
      544 GluAspSerLysGlnLysThrTrpCysHisLeuGlnGlnCysAsnArgLeu 563
      4327 GAACATATAGAAAGCTGGAATTCCAATGTGCTTCTTGAAAG----- 4371
      564 G1YThrThrGluLysLeuValLeuLeuAspGlnLysLysAspThrLeuLysAsp 583
      4372 -----ACAAAGCAGAGCTACAGATGAGTGAAGAGACCTCATGATGAT 4416
      584 GlnGlu-----SerCysValLysTrpLeuLysGluLysCysAsn----- 596
      4417 GTGGAAGATCTATAGCTGCTGCTGATGCTGCTGATTAAGAAAGAACTTTGACAG 4476
      597 -----LysTrpSerArgArgGlyAspAspArgPheSerPheValCysValPheGln 613
      4477 GTTCTGCGAGATGAAACAGAG-----TATGAG 4506
      614 AsnAlaThrCysGluLysMetValLysAspValLysAspArg-----CysGluVal 630
      4507 GAAACTGAGCGCTGAACCTTGAGCGCTCCAGAAAGAGTCGCTTCTTCAGACATGAGCTG 4566
      631 PheLysLysAsnLysLysLysSerTrpLysLeuPheLeuGluAsnAsnThrAsnLys 650
      4567 TTCAAG-----GTGAAGAAAGCTTAC-----GAGGAATCCCTGATCAT 4605
      651 IleThrThrLeuGluArgAsnCysProSerTrpHisLeuThrCysAsnArgPheSerPro 670
      4606 CTGAAACTCTTAAGCGAGAG-----AATAAGAACTTCAACACAG 4644
      671 AsnCysProGlyLeuThrLysGlu-----AsnSerCysThr 682
      4645 GAGATTTCTGACCTGACGAGCAAAATTGACAGAGGCTGAAACATATTCATGAATCTGAG 4704
      683 LysLysLysLys----- 686
      4705 AAAGTAAAGAAACAACCTTATCATGAGAAAGTGAACATCAAGACTTCCCTAGAGAAACA 4764
      687 -----HisArgGluProPheTrpLysArgLysAlaLeuGlu--AspAla 700
      4765 GAGCATCTCTTGAAGATGAAGAGCAAAATTCTTCGATTTCAACTGAGACTAAATGAG 4824
      701 LeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCysGluProAlaLeuLys 720
      4825 GTGAATCTGATGATGACCGAAAAATGCTGAAAAAGATGAAGAACTGCATGACTAAAG 4884
      721 Arg----- 721

```

```

Db      4885 AGGAACCATCTCAGAGTTGTGAGTCAATGACAGATCACTGATGCTGATCAGAGC 4944
      722 -----TyrCysThrValAlaGlyAsnValAsnAsnAlaSerIle 734
      4945 AGAAATGATCTCTGACGATCAAGAAAGAGATGAGAGGAGATCTTAATGAATGAAATC 5004
      735 SerGlyLeuCysValAlaAsnThrLys----- 743
      5005 CAG---CTGAACCATGCGAACCGCAGGCTGTGAGGCACTAAGAAATCTTAGAAACACA 5061
      744 -----AspAsnSerGlyLysSerAspGluAsp 752
      5062 CAAGAAATCTGAAGAGCACTACAGTCAATCTTGATGATGATCCATCAAGGCCAAGATGAC 5121
      753 AlaArgLysGluLeuCysGluLysLeuValLysGluValGlnGlnCysLysAlaLeu 772
      5122 CTTAAGGAACAAATTGGCA-----ATGTTAGCGCAGAGCTAACCCTGATG 5166
      773 ProThrGluLeuGlnProAlaAlaAspLeuLysAspTrpLysThrTrpGluGlu 792
      5167 CAGGCTGAAGTTGAAGAGCTCAGGGCATCCCTGGAACGACT-----GAGAGA 5214
      793 LeuLysLysArgAlaGlnGluAlaMetAsnLysSerSerLeuValLeuSerLeuIle--- 811
      5215 GGCAGAAAAATGCGACAGCAAGAGCTTCTGATGCCAGTGAACGTGCAACTTCTGCAC 5274
      812 -----LysLysAsnGluSerAsnValSerLys--- 820
      5275 ACTCAGAACCAAGCCTGATCAACCAAGAAAGAGGTGAACAGACATTTCCCAATC 5334
      821 -----SerAsnSerLysAsnLysAspLysAsn 829
      5335 CAGGAGAGATGAGAGCATGCTCCAGAAAGCCCGCAATCAAGAGAAAGGCCAAGAG 5394
      830 AlaValSerAsnGlyLeu-----GlnAspThrLys 840
      5395 GCCATCACTAGTGTGCTCATGATGCTGAGAGAGCTGAAGAAAGAAAGGACACAGCGGCC 5454
      841 HisValValLysLeuArgArgGly-----ValLysAspVal----- 852
      5455 CACCTGAGGCGATGAAGAAAGAACATGAGACAGACCTGAAGATCTGACGCTCGCTG 5514
      853 -----SerValThrGluLeuGluAla 859
      5515 GGTGAGCTGAGCAGCTGCGGCTGAAGGTTGGGAAGAAAGCATTCGAAGAACTGAGAGCC 5574
      860 LysAlaPheAspLeuAlaGlnValPheGlyArgTrpValAsp----- 874
      5575 AGGCTGAGAGAGCTTGAAGATGAGGTGAAGTGAACAGAAACCAATGTTGAGGCTGTC 5634
      875 -----LeuLysGluArgCysAsnLysLysGluSerAsp 885
      5635 AAGGCTTTCGCAACATGAGAGAGAGTGAAGTGAAGTCACTTACAGACTGAGAGAGAC 5694
      886 -CysArgLysLysGluAspCysLysAspLeuGlnGluValCysLysLysLysLeuAsnLysAl 905
      5695 CGCAAGAAATATCTCAGAGCTGACAGACTTGTGACAAATTTGCCAAACCAAGTCAAGCT 5754
      905 aCysArgAsn--LeuLysProLeuGlnValLysProHisGluThrValThrGluSerTh 924
      5755 TACAAAGACAAAGTGAAGAGGCTGAGAAACAAATCCAAATGCAACTTGCACAGTTCCGC 5814
      924 rThrThrThrThr-----ThrThrThrValAlaAspProLysAl 939
      5815 AAGCTTCAGACGAGCTGAGAGAGCGGAGAGACGAGTCAATGCTGAGTCCCAAGTC 5874
      939 aThr 940
      5875 AACA 5878

```

RESULT 14  
 US-09-949-016-2264  
 ; Sequence 2264, Application US/09949016







```

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4206
; TYPE: DNA
; ORGANISM: Bos taurus
US-09-750-590A-3

Alignment Scores:
Pred. No.: 2,936-17 Length: 4206
Score: 262.00 Matches: 226
Percent Similarity: 34.94% Conservative: 151
Best Local Similarity: 20.95% Mismatches: 380
Query Match: 4.84% Indels: 323
DB: Gaps: 46

US-10-654-416-14 (1-1023) x US-09-750-590A-3 (1-4206)

QY 5 VallyysarGlnAlaAlGlyThrGln-----AenSerIleaspGluGlu 19
DB 1606 GTGAAATACGAAGAGTGCCTCCGACAGAGTGGGAAATTGAGAAACCAATCAAAAT 1665
QY 20 HisValleuAlaLeuIleLeuLyseGluAapGlyLeuSerGluGluGlnGlyCysLyblyblys 39
DB 1666 GAATGCTTAACTTGAAGAGATTAAAGAGATGACGGCAAGCTGATGAGAAAGAAATTAAGCA 1725
QY 40 LeuLyblybTyrcyGgInGluLeuThrGluAlaLybLeuAenIleGluGlnValHisArg 59
DB 1726 CTGCAGAG-----GAGTTGACATGCTGTGAATCGAGCCGAGAGAGAGAGAGAGG 1776
QY 60 LybLeuLybGlyPheCysGluAapGlyblybAlaAapThrLybCysLybGluLeuLybAla 79
DB 1777 AAGCTCACT-----GAGATGAGAGGCCAGTTTAAAGACTTGTCAACC 1818
QY 80 AenIle-----GluLyblysCysThrIleLybGly 90
DB 1819 AAGCTGGCCCTTCTATTCACACAGAGAAATTTGAAAACATGAAGACTTGTATCAAT 1878
QY 91 LybLeuLybGluAlaIleLybLyblybIleGlnIleIleThrAapLybAapCysLybGlu 110
DB 1879 GAATGACGAGAGAGGCAAAAAATTAATGATGAGAAAGAGAAATATGAATGATCACTT 1938
QY 111 AenGluGlnGlnCysLeuPheLeuGluGlnGlyAlCysSerLybGluLeuLybAapCys 130
DB 1939 AATGAA-----ACTAGACCATTAAGAGAAACTT 1968
QY 131 AenThrLeuArgAenLybCysTyrcGlnLybLybAapLybValAlaGluGluValLeu 150
DB 1969 GAGAAATTTGAAGGCCAACTGGCTCAGACCTCAACCAAGAGAAACATGACAGCTCAG 2028
QY 151 LeuArg--AlaLeuArgSerAapLeuAenGlySerValIleCysGlu----- 165
DB 2029 AGCAGATTAGACGAAGTCTGAGAGAACTTGGGAAAGAGATCACTGATTAACATCGAAA 2088
QY 166 -----LybLeuLeuLybGluIleCysProValMetGlyArgGluSerAap 180
DB 2089 AATCAGACGTTACAAAGAAATCGAAAAGCTCTCGATTAATTAACCTCTTACAAA 2148
QY 181 GluLeuThrAenLeuCysLeuAenGlnLybGluThrCysLybAenIleLeuIleGluLyb 200
DB 2149 CAAGTAATTAATCTTAACAACGAAATGAAAATGCCCTTTAAAAGTAAAGTGAAGAAATG 2208
QY 201 AapLybLybCysGlyThrLeuLybThrAapValSerAlaIleLeuGlySerPheLyb--- 219
DB 2209 AAAAAGTCACATGATGATTAATGTTGATTAATAAAGCTTTCAGATGTGACACAC 2268
QY 220 -----LybGluThrCysLeuGlnLeuLeuGlnGlnCysTyrcPheThrIleGlyAenCys 237
DB 2269 AAATATACAGAAAGAACGTTGAAATGGA-GAAGTTGCTTAT-----GGAAGAAAGC 2318
QY 238 GLybAapAapIleIleLybCysIleGluLeuGlyLybGlnGluGlnAlaAenIle 257
DB 2319 -----CAGTTTAAGTAAATAATGTCAG----- 2339

```

```

QY 258 AlaTyrcMetProPheGly-ProAapPheAapProThrArgProGluAlaThrIleAlaGly 277
DB 2340 -----CCGCTCGAAACTGTGTTCAATACCT-----CCCAAGACACGAAAAAGA 2384
QY 277 uAapIleGlyLeuGlu-----GluPheTyrcLybLybValGluGluAapGlyb 293
DB 2385 AATGATGCTCTGAAATCCAAATATCACTGAACTTAAGAACGAGCTGTGAA----- 2436
QY 293 lPheIleGlyLybAenHisLeuArgAapAlaThrAlaLeuLeuAlaLeuLeuIleGlnAs 313
DB 2437 -----CTTAATTAATAAATGTGTGTAAGCAAGAGAAATATATTCACATCATGTCTGAAA 2492
QY 313 pSerSerLeuLybLyblys----- 319
DB 2493 CAATGATTTGAAAAAGACATGACTCATGATATGTCCTCGTGAAGAAACCCATGAAGAGAT 2552
QY 320 -----AapAapLybGly 323
DB 2553 TAAATCTGCTTGATGACACATTTGATTAATAAACAATGACAAATTAAGATGTGAAGAA 2612
QY 323 uLybCysGluGluAlaLeuGlnLybSerCysLybAenProHisGluHisGluAlaLeuGly 343
DB 2613 GAATGTTGAAGATTAATCAAGAAATTTGTGAAAATTAAGATGAAACGAAATATTA-- 2670
QY 343 uSerLeuCysLybLybAenGlyLeuSerAapAap----- 354
DB 2671 -----AAAAAATCTGGAGAAACATCAGAACCCAACTAAAGCTGATCAT 2717
QY 355 -GlyThrLybLybCysGlyGluGlnGlnAenAapIleAenLybThrCysLybIlePheThr 374
DB 2718 CAGCTTAAGAGAGCATGAAGAAAGATGAGTGCCTTAAGAGAGCATGAAG----- 2769
QY 374 rSerLybValThrAenAenArg-----LeuPheAapProThrLybGlyAenAenGly 391
DB 2770 -----AAGTCCAGGACAAACAGCCGTGAATATCTGCTAAGTACAAAACCAAGAGAGA 2825
QY 391 uIleValGlyTrp-----GluGlyLeuProThrPheLeuSe 403
DB 2826 GATTGTCAACCTCGATGAGAGATTTGACGCCACAGAAAGAGAACTTCGACACAGATA-- 2880
QY 403 rAenGluAapCysAlaLybLybLeuGlnuserTyrcPheTyrcPheGluLybLybCysProAs 423
DB 2881 -----CAGAAATGATCAAGGCTAAATATGCTCCGATCATCACTTGGAAGGTGAG-- 2934
QY 423 pGlyGluAenAlaCysLybAenIleArgAlaThrCysTyrcLybArgGlyLeuAapAlaAr 443
DB 2935 -----AGAAAATTTAAAGCACCT----- 2952
QY 443 gAlaAenLybValLeuGlnGlnAenMetArgGlyMetLeuHisGlySerAenLybSerTr 463
DB 2953 -----GAGAAAGAACTTAAGAAAGACGCTATCCACAGACACAGAAATATTAATACCACT-- 3006
QY 463 pLeuGlnLybPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 479
DB 3007 -----GAAAGAGAGGCCAAGAGAGTGAAGATTTTCAATTTGAGAAATTC 3050
QY 479 sLybGluAenLybGlySerPheSerAenAapGluLeuPheIleLeuCysValAlaGlnProAl 499
DB 3051 GAAGAG-----ATCTTCACTCTTCAG----- 3072
QY 499 aLybAlaIlaArgLeuLeuThrHisAapLeuArgMetLybThrIlePheLeuArgGlnGly 519
DB 3073 -----AAGATCTTAAGAGATTAAGATTTTCAATTTGAGAAATTC 3110
QY 519 nLeuAapGlnLybArgAapPheProThrAapLybAenCysLybGluGlnGlnGlnGlnGlnGln 539
DB 3111 TTATGACGAGAAAGAA-----GCATTAAGCAGAAAAACAGAAAGCTGAACACAGCTT 3164
QY 539 sGlnAapLeuGlnGlyGluAapSerLybGluIleThrTrpProCysHisThrLeuGlnGlnGly 559
DB 3165 AAAAGACCTGTTGCAAGAAATACACAGAGGCAAAAGAGAAAGAAAGAAAGCTTCGTGAGGA 3224

```

Qy 559 nCyAaenArlgLeuGlyThrThrgluLeu---LySglValleuLeuAerGluHlVly 578  
Db 3225 A---AATGCCAAGCAGACGACTTGTGAGATCTTGAGACCAAACTCTTTGAGAAAGCAGCA 3281  
Qy 578 aAerThrlEuLyAerGlnGluSerCyAvalLyerThrlEuLySglVlyCyAenLyStr 598  
Db 3282 TGTTCGGGTGAGCAGGTGAGTGCCTGAAAAATCTTT----- 3321  
Qy 598 pSerAArgAglYlAerAerAerGpHeSerPheValCyvalPheGlnAenAlaThrCySgl 618  
Db 3322 -----AGTGGTACAAATGA 3335  
Qy 618 uLeuAerValLyAerValLyAerAerGpCySglValPheLyAerAenLySgl---Al 637  
Db 3336 GACACTCAAGAGAACTGAATAAATGACAGAGATGTATAGAAAAGAGCAGCAGCAGACT 3395  
Qy 637 aSerTrtleleGluPheleGluAenAenThAenAenThAenAenThAenAenThAenAen 657  
Db 3396 GACCCAACTGGCGCAGATGTGAGAAAT----- 3423  
Qy 657 nCyAerAerThrlAerThrlThrThrThrThrThrThrThrThrThrThrThrThrThr 677  
Db 3424 -----CA 3425  
Qy 677 aGluAerSerCyAerThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 697  
Db 3426 GAAGAACTCTGTGAGCCCTGGGCTGAGCATTTGACG-----GT 3464  
Qy 697 uGluAerAlaLeuLyAenValGluLeuGlnGlyLeuLeuThAerLySerLyCySglPr 717  
Db 3465 TAAAGAAACATTTGAGAAAGAAAT----- 3489  
Qy 717 oAlaLeuLyAerGlyThrThrValAlaGlyAenValAenAenAlaSerIleSerGlyLe 737  
Db 3490 -----GGAAAT 3494  
Qy 737 uCyAerValAaenThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 756  
Db 3495 CATAAAGCTGCTTGAGAGAA-----AAGGAAAGAAAGCCAAACAAACTGA 3545  
Qy 756 uLeuCySglYlAerValLySglValGluGluGlnCyLyAerValAenProthrgluLe 776  
Db 3546 AGAGGTCTCCAAACTCCAGTGTGAGATTCAGAAATCTAAACAGCGTTAAAAAATTGA 3605  
Qy 776 uGlyGlnProthrlAaerLeuLyAerThrlThrlThrlThrlThrlThrlThrlThrl 796  
Db 3606 GACTCGGAGGTGCTGTATTGTGAAA---TATTAAGCA----- 3642  
Qy 796 gAlaGluGluAlaAerAenLySerSerLeuValleuSerleuIleLyAerAenGluSe 816  
Db 3643 -----ACGAAAGCGATTG-----GAGAC 3662  
Qy 816 rAenValSerLySerAenSerLyAenLyAerLyAenAlaValSerAenGlyLeuG 836  
Db 3663 ACAGATTCTCCAGTTAAAGAAATTTGGCAATCTGAATAGCAAGTATGAGCAAGTATG 3722  
Qy 836 nAerThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 856  
Db 3723 TGAGAGGTGTTGATGCTCCAAAGAAAGAAAGAACTGTGCTAAAGAT-----GA 3770  
Qy 856 uLeuGluAlaLyAerValPheAerLeuAlaGluValPheGlyAerGlyValAerLeuLy 876  
Db 3771 GAAGGAATGTGCTCATTTCAAGATAGAGCAAGAAATC-----AAAGATCAGCA 3818  
Qy 876 aGluAerGpCyAerAenLySgl-----LeuGluSerAerCyAerGlnIleLySglA 891  
Db 3819 GGAACGATGTGACAAATCTTAACAAACATCAGGAGCTACAGAGAAATACAGAAATC 3878  
Qy 891 pCyAerLyAerLeuGluGluValCyLyAerValleuAenLyAerAerAenLeuLyPr 911  
Db 3879 TGCCAAACAAATCGAAAGAAAGATTAATGAATTAAGTGAATGCTCAATGATGTGGAGAG 3938  
Qy 911 oLeuGlu-----ValLyAerProthrgluThrValThrgluSerThrlThrlThrlTh 929

Db 3939 ATTAAGAGGCGCCCTCAATGCGCTTTCCAGCTCAGCTCACTAATGAAGTGAAGTCCAGCA 3998  
Qy 929 rThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 949  
Db 3999 GAGCAGAGTCACTGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4054  
Qy 949 aThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 968  
Db 4055 --ACGTGGCGGATGCCAGACAGACAGCAAGAAATGATTCATTTATGAGACAACT 4112  
Qy 968 lAerSerLySglThrlThrlThrlThrlThrlThrlThrlThrlThrlThrlThrl 985  
Db 4113 TCTTAAGTGTGACAGGCTCAGATGATGAGATGATGAGATGATGAGATGATGAGATG 4165

Search completed: October 29, 2005, 15:19:52  
Job time : 573 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_blue\_p2n model

Run on: October 29, 2005, 09:48:47 ; Search time 1797 seconds  
(without alignments)  
4701.696 Million cell updates/sec

Title: US-10-654-416-14  
Perfect score: 5411  
Sequence: 1 MARAAVKRQAAGTQNSIDEH.....MSGNMVRGVAVMVISFMI 1023

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9784742 seqs, 4129495052 residues  
Total number of hits satisfying chosen parameters: 19569484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh  
-O=/cgnt\_1/USPTO.spool/US10654416/funcat\_26102005\_101025\_21801/app\_query.faeca\_1.1159  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.tmpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62  
-TRANS=human40.cdd -LIST=45 -DOCALL=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=0 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10654416@CGN\_1.1159@funcat\_26102005\_101025\_21801  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.\*  
1: /cgnt\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgnt\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgnt\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgnt\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgnt\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgnt\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgnt\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgnt\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgnt\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgnt\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgnt\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgnt\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgnt\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgnt\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgnt\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgnt\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
17: /cgnt\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
18: /cgnt\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgnt\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
20: /cgnt\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
21: /cgnt\_6/ptodata/1/pubpna/US10H\_PUBCOMB.seq.\*  
22: /cgnt\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
23: /cgnt\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
24: /cgnt\_6/ptodata/1/pubpna/US10I\_PUBCOMB.seq.\*  
25: /cgnt\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*  
26: /cgnt\_6/ptodata/1/pubpna/US11A\_PUBCOMB.seq.\*  
27: /cgnt\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
28: /cgnt\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5411	100.0	3072	19 US-10-654-416-13	Sequence 13, App1
2	4398.5	81.3	3054	19 US-10-654-416-11	Sequence 11, App1
3	4352	80.4	3084	19 US-10-654-416-7	Sequence 7, App1
4	3827	70.7	3090	19 US-10-654-416-5	Sequence 5, App1
5	3401.5	62.9	3081	19 US-10-654-416-9	Sequence 9, App1
6	2752.5	51.1	3042	19 US-10-654-416-1	Sequence 1, App1
7	2404.5	44.4	3006	19 US-10-654-416-3	Sequence 3, App1
8	396	7.3	249	19 US-10-654-416-15	Sequence 15, App1
9	331.5	6.1	5784	18 US-10-282-122A-22545	Sequence 22545, A
10	312.5	5.8	5460	19 US-10-335-977-3230	Sequence 3219, Ap
11	311.5	5.8	5460	19 US-10-335-977-3230	Sequence 3220, Ap
12	308.5	5.7	5334	10 US-09-882-227-623	Sequence 623, App
13	274	5.1	5373	18 US-10-369-493-25273	Sequence 25273, A
14	274	5.1	5641	17 US-10-032-585-6646	Sequence 6646, Ap
15	273.5	5.1	3492	18 US-10-282-122A-15680	Sequence 15680, A
16	270.5	5.0	4087	18 US-10-369-493-26896	Sequence 26896, A
17	270	5.0	5886	21 US-10-798-037-3	Sequence 3, App1
18	270	5.0	6010	21 US-10-723-860-57	Sequence 57, App1
19	270	5.0	6010	21 US-10-723-860-105	Sequence 105, App
20	270	5.0	6010	21 US-10-723-860-4858	Sequence 4858, App
21	270	5.0	12801	21 US-10-798-037-5	Sequence 5, App1
22	269.5	5.0	4780	9 US-09-962-436-287	Sequence 287, App
23	269.5	5.0	4780	22 US-10-843-641A-2746	Sequence 2746, Ap
24	267.5	4.9	7985	26 US-11-097-143-9827	Sequence 9827, Ap
25	266	4.9	6965	21 US-10-335-053-229	Sequence 229, App
26	266	4.9	7694	17 US-10-096-534-34	Sequence 34, App1
27	266	4.9	7695	11 US-09-968-007A-462	Sequence 462, App
28	266	4.9	7695	22 US-10-843-641A-6932	Sequence 6932, Ap
29	266	4.9	8063	10 US-09-814-353-21776	Sequence 21776, A
30	261	4.8	5835	10 US-09-927-597-1	Sequence 1, App1
31	261	4.8	5835	24 US-10-486-057-1	Sequence 1, App1
32	261	4.8	5937	10 US-09-927-597-3	Sequence 3, App1
33	261	4.8	5937	24 US-10-486-057-3	Sequence 3, App1
34	261	4.8	6861	15 US-10-171-311-161	Sequence 161, App
35	261	4.8	6861	24 US-10-341-434-102	Sequence 102, App
36	261	4.8	6900	15 US-10-923-035-18	Sequence 18, App
37	261	4.8	6900	15 US-10-171-311-163	Sequence 163, App
38	261	4.8	6900	24 US-10-764-425-13	Sequence 13, App1
39	260.5	4.8	5780	19 US-10-336-472-19	Sequence 19, App1
40	260.5	4.8	6773	9 US-09-864-864-136	Sequence 336, App
41	260	4.8	4428	18 US-10-395-027-445	Sequence 445, App
42	258.5	4.8	3570	19 US-10-336-472-15	Sequence 15, App1
43	258	4.8	11065	15 US-10-116-802-14	Sequence 14, App1
44	256	4.7	4643	9 US-09-864-864-266	Sequence 266, App
45	256	4.7	5956	24 US-10-783-271-100	Sequence 100, App

#### ALIGNMENTS

RESULT 1  
US-10-654-416-13  
; Sequence 13, Application US/10654416  
; Publication No. US20040043409A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/10/654,416  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US/09/762,724  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750

PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: US 60/096,805  
PRIOR FILING DATE: 1998-08-17  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 3072  
TYPE: DNA  
ORGANISM: Pneumocystis carinii sp. f. hominis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) .. (3072)  
US-10-654-416-13

Alignment Scores:  
Pred. No.: 0 Length: 3072  
Score: 5411.00 Matches: 1023  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
DB: 19

US-10-654-416-14 (1-1023) x US-10-654-416-13 (1-3072)

QY 1 MetAlaArgAlaValLysArgGlnAlaLeuThrGlnAsnSerTLeuAspGluLys 20  
DB 1 ATGGCGCGCGCGTACGCGCGAGCGAGGACAGAAATGCAATGATGAGAGAAAT 60  
QY 21 ValLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
DB 61 GTTTTAGCTTTAAATCTTAAGAGAAATGAGCTTAATGACAGAAATGCAAAAAA 120  
QY 41 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60  
DB 121 AAAAAATTTGCGCAAGAAATTCGACGACAAACTAAATATGAAACAAGTACACAGAAA 180  
QY 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80  
DB 181 CTTAAAGGTTTTTGGCAAGATGAAAGACAGATACAAATGCAAAATGCAAAAGCTGAAGCCAA 240  
QY 81 LLeuLysLysCysThrThrLysGlyLysLysGluAlaLeuLysLysLysLysLys 100  
DB 241 ATTGAGAAATAATGATCAATCAATCAAAAGAAATCTTAAGAAAGCAATTAATAAAAAA 300  
QY 101 GlnLeuLeuThrAspLysAspCysLysGluLeuGlnGlnGlnGlnGlnGlnGln 120  
DB 301 CAGATTAAACGATTAAGATTCGAAAGAAATGAAACAATGCTATTGTTGAGGGA 360  
QY 121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysThrGlnLys 140  
DB 361 GATGTTCAAAAGAACTTAAGATGATGCAATTAATTTGAGAAATTAAGTCTATCAAAAAG 420  
QY 141 LysArgAspLysValAlaGlnGluValLeuLeuAlaLeuArgSerAspLeuAsnGly 160  
DB 421 AAAGCTGATTAAGTTGCGGAGAAATCTTTTAAGAGCACTTCGAGGATCTTTAAATGGA 480  
QY 161 SerValLysCysGluLysLysLysGluLysCysProValMetGlyArgGlnSerAsp 180  
DB 481 TCAGTCATGATGAAAAAACTTAAGATTCCTGCTGCAAGGAGGAGGAAAGATGAT 540  
QY 181 GlnLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnLysLeuLysLys 200  
DB 541 GAGTTAAACAACTGTTGTTGTAACCAAGAAAGAGATGATTAAGAAATTAATTAAGAAAA 600  
QY 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaAlaLeuGlySerPheLysLys 220  
DB 601 GATAAGAAAGTGGGATCTTTAAACAGATGTTTACGACGACTGAGAAATTTAAAAAA 660  
QY 221 GlnThrCysLeuGlnLeuLeuGlnGlnCysThrPheThrLysLysLysLysLysLys 240  
DB 661 GAAACATGCTTTGATTAAGTACGAAACAATGCTATTTTACATTTGAAATTTGCGAGACAC 720  
QY 241 AspLeuLeuLysCysLeuLeuLeuGlnGlnLysCysGlnGlnGlnGlnGlnGln 260

DB 721 GATTAATTAATGATTTGAAATGCGAGAAAAAGCCCAAGAAACAATAATGCTTAATG 780  
QY 261 ProProGlyProAspPheAspProThrArgProGluAlaThrLysAlaGluAspLys 280  
DB 781 CCACAGGACCCGATTTTGAATCCAACTAGGCAAGGCTTACATAGCAGAGATTAAGG 840  
QY 281 LeuGlnGluPheThrLysLysValGlnGluAspGlyValPheLysLysLysLysLys 300  
DB 841 CTGAGAGGTTTTTAAGAAAGTAGAGAGATGAGATTTTAATTTGGAACAAATCATCTA 900  
QY 301 ArgAspAlaThrAlaLeuLeuAlaLeuLeuLysGlnAspSerSerLeuLysLysAsp 320  
DB 901 AGAGATGCGACAGCTTTGTTGCAATTTGATTCAGATTCATTTAAAAAAAGAAC 960  
QY 321 AspLysGlnLysCysGlnGluAlaLeuGlnLysSerCysLysAspProLysGlnGlu 340  
DB 961 GACAAAGAAATTCGAAAGAAAGCCCTTCBAAAAAGCTGCAAAAAATCCCAAGAACATAG 1020  
QY 341 AlaLeuGlnSerLeuCysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGlu 360  
DB 1021 GCTTTAGAAATTTATGTAAGAAAAATGCTTTAATGATGAGAACCAAAAAATGTGAA 1080  
QY 361 GlnLeuGlnAsnAspLysAsnLysThrCysLysLysLysPheThrSerLysValThrAsnAsn 380  
DB 1081 GAATTCGCAAAATGATTAACAAACTTCGCAAAATTTCACTTCBAAAAAGTCACTAATAT 1140  
QY 381 ArgLeuPheAspProThrLysGlyAsnAsnGlnLysValGlyThrGlnGlyLeuProThr 400  
DB 1141 GGTCTTTTGTATCCACAAAGAAATTAATGAAATTTGTGATGAGAAAGGTTGCAACA 1200  
QY 401 PheLeuSerAsnGlnAspCysAlaLysLeuGlnSerThrCysPheThrPheGluLysLys 420  
DB 1201 TTTCTTAGCAACGAAGATTTGCGCAAAATTCGAGTCCATATGCTTTATTTTGAAGAAAA 1260  
QY 421 CysProAspGlyGluAsnAlaCysLysAsnLysArgAlaThrCysThrLysArgGlyLeu 440  
DB 1261 TGTCCAGTGGAGAAATGCTGTAATAATTAAGACACATGTTACAAAGAGGACTT 1320  
QY 441 AspAlaArgAlaAsnLysValLeuGlnGlnLysAsnMetArgLysMetLeuLysGlySerAsn 460  
DB 1321 GATCAGCGCGCAATTAAGTCTCTCAAGAAATTAAGGAGAAATGTTACATGCTTCAAC 1380  
QY 461 LysSerThrLeuGlnLysPheGlnGlnGluLeuValLysValCysGluLysLysLys 480  
DB 1381 AAAAGCTGGCTGAAAGTTTCAACAAAGATTAATAAGATATGAGAACTGAAAAA 1440  
QY 481 GlnAsnLysGlySerPheSerAsnAspGluLeuPheLysLysLysLysLysLysLys 500  
DB 1441 GAAACCAAGAAATTTCTCAACAGATGATTAATTTATCTGTGTATCACACAGCAAAA 1500  
QY 501 AlaAlaArgLeuLeuThrLysAspLeuArgMetLysThrLysPheLeuArgGlnGlnLeu 520  
DB 1501 GCACCCCGGTTGCTTACACATGATCTTCGAATGAAAACTATCTTTTAACGACAACTG 1560  
QY 521 AspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGln 540  
DB 1561 GATCAAAAGCGAGATTTCCCAAGATTAATAATGCAAGAAATTTGGGAGGAAAGTCCAA 1620  
QY 541 AspLeuGlnGluAspSerLysGlnLysThrThrProCysHsThrLeuGlnGlnGln 560  
DB 1621 GATTTAGAGAGATTCACAAAGAAATTAACATGATGCTTAACATGAGACACCAATGC 1680  
QY 561 AsnArgLeuGlyThrThrGlnLysLeuLysGlnAlaLeuLeuAspGlnLysLysAspThr 580  
DB 1681 AATGCTTTGGGATCAACAGAAATTTTAAGAGAGTTTATTTGATGAAACAAAGATCT 1740  
QY 581 LeuLysAspGlnGlnSerCysValLysThrLysLysGlnLysCysAsnLysThrSerArg 600  
DB 1741 TTGAAGACCAAGAAATTTGTGTAATAATCTTAATAAGAAAGTGAATTAATGCTCTGA 1800  
QY 601 ArgGlyAspAspAspPheSerPheValCysValPheGlnAsnAlaThrCysGlnLeuMet 620





Db	301	ACAAATCCA-----TCAGATGATTAATTCGAAAGAGAGTGAACGACATGGCTTA	348
Oy	117	PheLeuGluGlyValCysSerIleGluLeuIleuIysAspArgCysAsnThrLeuArgAsnIle	136
Db	349	TTTTTGGAGGGGAGCATGCG---CATATCTGTGTAAGATTTGTAACAAACTAAGAACTTA	405
Oy	137	CysIlyrGlnIleValArgAspIlyValIleGluGluValLeuLeuArgAlaLeuArgSer	156
Db	406	TGTTACCGAAAAAAACGTGACCGAGTACAGAGAAAGTCTTTTGAGGGCACTTCGATG	465
Oy	157	AspLeuAsnGlySerIleIleCysGluIlyValIleGluIleCysProValMetGly	176
Db	466	GATCTCAATTAATAACGAAACACATGAAATAAAACGAAAGATTTGGCCAGTCTTCGAC	525
Oy	177	ArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnIlyGluThrCysIleAsnIle	196
Db	526	AGCGAAAGTAAGATTAACCGACCTTGTTGTTGAACCGAAAAAGACCTGGCGAAATAT	585
Oy	197	LeuIleGluValArgIlyValCysGluIlyThrLeuIlySerThrAspValSerAlaIleGly	216
Db	586	ATTAATAAGAAAGATTAATAAAATGACACTCTTTAAAGCAAAATGGTGCACAGCACTTGA	645
Oy	217	SerPheIlyValGluThrCysLeuGluLeuLeuGluGlnCysIlyrPheIlyrIleGlyAsn	236
Db	646	AGTTTAAABAAAAGAAATATGCTTGAATTAATCTGAACAATGCTATTTTAACTATGGAAAT	705
Oy	237	CysGlyAspAspAspIleIleIleCysIleGluLeuGlyGlyIlyValCysGluGluGlnAsn	256
Db	706	TGCGGAGACGACGATATTAATTAATGTATTTGAATTGGAGGGAATATGCCAAGAACAAAC	765
Oy	257	IleAlaIlyrMetProProGlyProAspPheAspProThrArgProGluAlaThrIleAla	276
Db	766	ATTGTTTAAATACCAACGAGGACCCGATTTTGATCACTGAACCAAGAGGCTACACTGACA	825
Oy	277	GluAspIleGlyLeuGluGluIlyrPheIlyrIlyValIleGluIleAspGlyValPheIleGly	296
Db	826	GAGGACATAGACCTGATGATGACTTTATTAATAAGCAGAGAGAGATGGTCTTTTATTCGA	885
Oy	297	IlyAsnIleLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIle--GlnAspSerSer	315
Db	886	AAACATCATTTTAAAGATGCGACACCTTTATGACGTTGGTTGTTAAGAAAGATGATAC	945
Oy	316	LeuIlyValIlyValAspAspIlyGluIlyCysGluGluAlaLeuGlnIlySerCysIlyAsn	335
Db	946	GGAAAAATTAATTAATATATCGAGAAAAATGCAATTAAGATCTTCGAAGATTAATGCAAAAAC	1005
Oy	336	ProIleGluIleGluAlaLeuGluSerIleuCysIlyValAsnGlyLeuSerAsnAspGly	355
Db	1006	TCTCAACGACATGAAGCTCTTAATAAATTAATATTAATTAATATGCTTAATGCAATAGCA	1065
Oy	356	ThrIlyValCysGluGluLeuGlnAsnAspIleAsnIlyrThrCysIlyrIlePheThrSer	375
Db	1066	AAAGAAAAATGCAAGATTAAGAAAGATTAATAAAAAACATGCACAAACCTCAAAACCA	1125
Oy	376	IlyValIleThrAsnAsnArgLeuPheAspProThrIlyGlyIleAsnGluIleValGlyrThr	395
Db	1126	ACGATTTCTTAATAAACCATCTTATATATCA-----ATGATTAATATGTTGAAGTGG	1176
Oy	396	GluGlyLeuProThrPheLeuSerAsnGluAspCysAlaIlyValLeuGluSerIlyCysPhe	415
Db	1177	AGAAAACTGCCAACATTTCTTAATGAAGACTGTCGACAGATTTGGATCTTAATGTTTTT	1236
Oy	416	IlyrPheGluIlyValCysProAspGlyGluAsnAlaCysIlyAsnIleArgAlaThrCys	435
Db	1237	TACTACGAAABAAAGCTTGTCCAATGCCAAGAAAGCTGATGAATTTGAGGGCAGCGTGT	1296
Oy	436	IlyrIlyValArgGlyLeuAspAlaArgAlaAsnIlyValLeuGluGlnIleAsnMetArgGlyMet	455
Db	1297	TATAAGAGAGGGCTTGATGAGAGGGCAATAAAGCTGCAGAAATAATATCGTGGCTTA	1356
Oy	456	LeuHleGlySerAsnIlySerThrLeuGluIlyrPheGlnIlyrGlnIleValIlyValCys	475
Db	1357	TTACGTGTTCAAAATCAAGTGGCTTAAGAGGCTTCAACAAAGATTAGTAAAGATATGT	1416

QY	476	GIuIyBleuLeuYelYvGIuAenLYleGlySerPheSerzAenAerGIuLeuPheIIeLeuCyS	495
Db	1417	AAAAAGCTA---AAAGAAAAATAAAGGAAGTTTCCCAAAACGATGAAATATTGTTCTGTGT	1473
QY	496	ValGIenProAlaLYbAlaAlaArgLeuLeuThrIIaAPLeuAArgMeLySThrIIePhe	515
Db	1474	GTACAGCCAGGAAAAGCTGCACGATTACTTACACACGATCATCAATAGAGGTTATCTTT	1533
QY	516	LeuAArgInGIuLeuAerGIuLYaYzAerPheProThrAplybaenCYbLYeGIuLeu	535
Db	1534	TTAGGACAAACAATTGGATCAAAAAGAGATTTTCCGACAGTAAAGACTGCACAGAAATTA	1593
QY	536	GIuYArgLYeCYsGIuAenAerGIuGIuAenSerLYeGIuIeThrTPProCYbHIAthr	555
Db	1594	GGGAAAAAATCCAGAGATTATAGAAAGGATTCAAAAAGAAATTACATGGCCATGTCAATCG	1653
QY	556	LeuGIuGIuInCYbaenAArgLeuGIYThrThrGIuIeLeuLYbGIuValIeLeuAerP	575
Db	1654	CTGGAGACGAATGCATCGCTTGGGGCTACAGAAATTTTAAAGCAGGTTTATTGGAT	1713
QY	576	GIuHIIeLYbAerPThrLeuLYbAerGIuIuSerCYbValLYeTYrLeuLYbGIuLYeCYs	595
Db	1714	GAAACAAAGATACTTTGMAAGACCAAGAAAGTTGTGTAATATCCATAAGAAAGAGCT	1773
QY	596	AenLYeTPSerAArgAArgIYAerAerPheSerPheValCYbValPheGIuAenAla	615
Db	1774	AATTAATGTGTTTAAGAAGAGATGACCGTTTCTCTTTTGATGTGCTTCCAAACGCT	1833
QY	616	ThrCYbGIuLeuMeCYaLYbAerValLYbAerValLYbAerPheCYbGIuValPheLYbAenIIe	635
Db	1834	ACGTGTGAGCTGATGTGTAAAGACGTGAAAGACAGGTGTGAAGATTCAAAAAATATA	1893
QY	636	LYbAlaSerTYrIIeGIuPheLeuGIuAenAerThraenLYbIIeThrThrLeuGIu	655
Db	1894	AAAGCTCATATATATTATTGAATTTCTTGAATATATCAATTAATAATTAACAACCTGAA	1953
QY	656	ArgAenCYbPProSerTPHIAThrTYCYbAenAArgPheSerProAenCYbPProGIYLeu	675
Db	1954	AGAAATGTCCCTCTGGCACAAGATATTGCAATAGATTTTCACTAATTGTCCAGGCTTT	2013
QY	676	ThrLYbGIuAenSerCYbThrLYbAIIeLYbLYbIIaArgIuProPheTYrLYbAArgLYb	695
Db	2014	ACGAAGAAGATAGTTGTGACAAAATACAGAACGATGTGAGCGGTTCTATAAAAAAGAAAG	2073
QY	696	AlaLeuGIuAerPAlaLeuLYbValGIuLeuGIuLYbLeuThrAerLYbSerLYeCYs	715
Db	2074	GCGTTTGAAAGTCTCTCAAGTGAAGCTTCAAGAAAATTTGACTGATAATCTTAATGT	2133
QY	716	GIuPProAlaLeuLYbAArgTYrCYbThrCYaAlaGIuAenValAenAenAlaSerIIeSer	735
Db	2134	GAACTGTGATTGAAAAGATATTGTAKCAGTACGGGAAACGTAAATTAATGCGCTAATCGT	2193
QY	736	GIuLYeCYbLYbAlaAenThrLYbAerAenSerGIuLYbSerAerGIuAerPAlaArgLYb	755
Db	2194	GCGTTATGCCAAAGCTTACACCCAGAGATACTCTGAAAAGACTGTATGAGGATGCTAGAAAG	2253
QY	756	GIuLeuCYbGIuLYbLeuValLYeGIuValGIuGIuInCYbLYbAlaLeuProThrGIu	775
Db	2254	GAACTCTGTGAATTTAGTAAAGAAAGTGGAAAGAACAGTCCAAAGCATTTTCCAAACAGAA	2313
QY	776	LeuGIuGIuProAlaAlaAerPLeuLYbLYbAerTYrLYbThrTYrGIuGIuLeuLYbLYb	795
Db	2314	TTAGGACAAACCGCAGCTGATTTTAAAAAAGATTATTAAGACATATGAGAACTTAAGAAA	2373
QY	796	ArgAlaGIuGIuAlaAerAenLYbSerSerLeuValIeUerSerLeuIIeLYbLYbAenGIu	815
Db	2374	CGTGAGAGGAAGCAATGAAACAGTCCAGTCTTGTGTCATCCATTAAGAAAAAGAA	2433
QY	816	SerAenValIserLYbSerAenSerLYbAenLYbAerLYbAenAlaValIserAenGIYLeu	835
Db	2434	AGTATATGTATCAAAAGTAAATAGCAAAACAGGATTAAGATGCCGTTTCAAAACGAGCTT	2493

[illegible]

Query Match:	80.43%	Indel:	26
DB:	19	Gaps:	6
US-10-654-416-14 (1-1023) x US-10-654-416-7 (1-3084)			
QY	1	MetAlaArgAlaValIlyrAGlnAlaAgIYthrGlnAsnSerIleAspGluGlnHis	20
DB	1	ATGCGCGCGGCGGCTCMACGGCAGCAAAAGTGCACGAATACATTGATGAGAGCAT	60
QY	21	ValLeuAlaLeuIleLeuIyGluAspGlyLeuSerGluGluGlnCylbYsYsYsLeu	40
DB	61	GTTTAGCTTGATTTTAAATAAAATGATTAAGAAATACAAATGCAAAATCAACTGTTG	120
QY	41	LybYsYrYrCyGgInGluLeuThrGlnAlaYsLeuMenIleGluGlnValHisArgYs	60
DB	121	GAGAGATTTTCACAAACATTAAACAATGACAGATTTAAATCCAGAAAAAGTTCCAGAAAA	180
QY	61	LeuYsGlyPheCyGgIyAspGlyLeuAlaAspThrYsYsYsGluLeuYsAlaAsn	80
DB	181	TTAAAGATTTCTGTGATTAACGGAAACGAATGAAAATGTCAGATCTTAAAAACAA	240
QY	81	IleGluYsYsCySthrThrIleYsGlyLybLeuYsGluAlaIleYsYsYsIle	100
DB	241	GTCATCAAAAATGATTAATTAATTAACAGAAAACCTTCAACAGCTGCAGAAAAAATT	300
QY	101	GlnIleIleThrAspYsAspCyYsYsGluAsnGluGlnGlnCylbPheLeuGluGly	120
DB	301	TCAGATTTAACAGATGAGGATTTGCAAAAAGATTAACAACATCTCTATTTTGGAGGA	360
QY	121	ValCySerSerYsGluLeuYsAspAspCyAsnThrIleuArgAsnYsYsYrGlnYs	140
DB	361	GCATGCTCCACAGAACTTAAAGATGACGCACATTAATTAAGAAATTAACGTTATCAAAA	420
QY	141	LybAsrAspYsValaIaGluGlnValLeuLeuArgAlaLeuArgSerAspLeuAsnGly	160
DB	421	GAGCGAACAATGTGGCGAGAAAGTCTTTTGAAGGGGCTTCGTGGATCTCAATGAA	480
QY	161	SerValIleCyGgIuLybYsLeuYsGluIleCySProValMetGlyArgGluSerAsp	180
DB	481	ACAAAGCATGTGAAAAAACTGAAGAAGCTTTGCCGAAATTAAGAAAGAAAGCAT	540
QY	181	GluLeuThrAsnLeuCybYsAsnGlnYsGluThrCybYsAsnIleLeuIleGluYs	200
DB	541	GAATTTAACGGACTTGTCTTTATCAAAAAACAACATGCGTAAGCTTTGTAACAAAGCA	600
QY	201	AspYsYsCySgIYThrLeuYsThrAspValSerAlaIleuGlySerPheYs---	219
DB	601	AAAAGTAAATGTGATCTCTTGAAGAAAGACATTTGAAGAGCATTTAAGAAATGAATG	660
QY	220	LybGluThrCybYsLeuGluLeuLeuGluGlnCylbYrPheYrIleGlyAsnCyGgIyAsp	239
DB	661	CGAGAAAATGTCTACTTACTTACGAGCAATGTTACTTCAACAAGGAGAACGTGAGAGA	720
QY	240	AspAsp-----IleIleYsYsYs	245
DB	721	GACCAATCAAGTCCAAATAACTTAATATTAAGAATGCAAGAAATATGTACAGAGCT	780
QY	246	IleGluLeuGlyGlyYsCyGgInGluGlnAsnIleAlaYrMetProProGlyProAsp	265
DB	781	GATGATTTAGCAAGAAAGTGTGAAAGAAATATTTGTTTATATGATCCAGATCCGAT	840
QY	266	PheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheYr	285
DB	841	TTTCATCCAACTTAAGCCAGACCTTACACTACACAGAGGACATAGAGGCTGGAAGACTTAT	900
QY	286	LybYsYsValaGluGluAspGlyValPheIleGlyYsAsnHisLeuArgAspAlaThrAla	305
DB	901	AAGAGGCGAGAGGAGATGGAATTTTGTTCGAAAGCAACATGTTAAGAGATGCAACAGCT	960
QY	306	LeuLeuAlaLeuLeuIleGlnAspSerSerLeuYsYsYsYsAspAspYsGluYsYs	325
DB	961	TTGTTGGCACTACTT-----CTTAAGAAAACCTTAAAAAAGAAAGATGT	1005

OY	326	GlucLUA1aLeuGlnLysSerCysblybAsnProHieGluLysGluAlaLeuGluSerLeu	345
		.....	
Db	1006	ATAAAGCCCTTAATAAAAAAACCTGGCAAAACCCCTCATGAACAATGAGGGCCTTGAAAACTCA	1065
OY	346	CysLysLysAsnGlyLeuSerSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAsp	365
		.....	
Db	1066	TGTAAAGGAAATTAACCAAGTAGTGATGAAACGAAATAATGTGAATCTAGAAAAAGAT	1125
OY	366	LeaAnLysThrCysblybLephethrSerLysValThrAsnAsnArgLeuPheAspPro	385
		.....	
Db	1126	GTTAAACAAACCTGTGACAAAGCTTACATCAACAATCTTAATAAACCGCTTACCTTTCA	1185
OY	386	ThrlYsgIyAsnAsnGluLysValcLysTrpGluGlyLeuProThrPheLeuSerAsnGlu	405
Db	1186	CCTGATGGA-----ATTGCGGAAAGGGGAAATTAACCGACATTCCTTAGACATGAA	1236
OY	406	AspCysAlaLysLeuGlnSerThrCysbPheTrpPheGluLysLysCysProAspGlyGlu	425
Db	1237	GATTGTGCAAACTAGAACTTATTGCTTTATTTATTAAGAACTTGCCAGATGCAA	1296
OY	426	AsnAlaCysblybAsnLysArgAlaThrCysTrpLysArgGlyLeuAspAlaArgAlaAsn	445
Db	1297	GAAGCTGTATGAATGTAGAGGGCAGCGCTTACAAGAGAGGGCTTGAGTCAGCGGCAAC	1356
OY	446	LysValLysGlnGluAsnMetArgGlyMetLeuHieGlySerAsnLysSerTrpLeuGlu	465
Db	1357	AGTGCTGTGCAAAATAATATCGTGCGGTATTAACGTGCTCAAAATCAAAAGTGGCTTAAG	1416
OY	466	LysPheGlnGlnGluLeuValLysValcLysGluLysLeuLysLysValAsnLysGlySer	485
Db	1417	GAGTTTCAACAAAGATTAGTAATAATAGTAAGAGCTA---AAAGAAATTAAGAAAGT	1473
OY	486	PheSerAsnAspGluLeuPheHieLeuGlyCysValGlnProAlaLysValaAlaArgLeuLeu	505
Db	1474	TTCCCAAAAGATGAATAATTTGCTGTGTGTACAGCCAGCAAAAGCTGCACGATTACTT	1533
OY	506	ThrHieAspLeuArgMetLysThrLLePheLeuArgGlnGlnLeuAspGlnLysArgAsp	525
Db	1534	ACACACGATCATCAAAATAGAGGTTACTTTTTCACACAACAATTGGATCAAAAGAGCAT	1593
OY	526	PheProThrAspLysAsnCysblybGluLeuGlyLysArgLysArgGlyGluAsp	545
Db	1594	TTTCCGACAGATTAAGACTGCAGGACATTAAGGAAAAATCCCAAGATTGAGAAAGGAT	1653
OY	546	SerLysGlnLysThrTrpProCysHieThrLeuGlnGlnGlnCysAsnArgLeuGlyThr	565
Db	1654	TCAAAAGAAATTAACGTGCGCATGTATACATCGAGCGACATGAAACGCTTGGGACCT	1713
OY	566	ThrgLysLeuLysGlnValLeuLeuAspGlnLysLysAspThrLeuLysAspGlnGlu	585
Db	1714	ACAGAAATTTTAACAGAGTATTATTTGATGAAACAACAAGATACTTTGAAGAACAAGAA	1773
OY	586	SerCysValLysTrpLeuLysGlnLysCysAsnLysTrpSerArgArgGlyLysAspArg	605
Db	1774	AGTTGTGTAATATCACTAAAGAAAGGTATTAATAGTCTACAGAGGAGATGACCGT	1833
OY	606	PheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValLysAspValLys	625
Db	1834	TTCTCTTTTGTATGTGTTTCCAAAACCGTACGTGTGACGTGATGTAAAAGACTGAA	1893
OY	626	AspArgCysGluValPheLysLysAsnLysValaSerTrpLysLeuPheLeuGlu	645
Db	1894	GACAGGTGTGAAGTATCAAAAATAATTAAGAGCTTCATATATTAATGAATTTCTTGA	1953
OY	646	AsnAsnThrAsnLysLysLysThrLeuGluLysArgAsnCysProSerTrpHieThrTrpCys	665
Db	1954	AATAATATCAAAATTAATAATACACACTGGAAACAAATTTGCTCTGGCATACGATATGC	2013
OY	666	AsnArgPheSerProAsnCysProGlyLeuThrLysGluAsnSerCysThrLysLysLys	685
Db	2014	AATAGATTTTCACTTAATTTGCTCAGAGCCTTACGAAAGAGATAGTTGTACAAAATCCAG	2073
OY	686	LysHieAspArgLysProPheTrpLysArgLysValaLeuGluAspAlaLeuLysValcLysLeu	705

Db	2074	AAGCATTGTGACCGGTTCTATTAAAAAGAAAGCCCTGGAAAGATGCTCTCAAAAGTACACTT	2133
QY	706	GIInGIlybSeuThrAapLySeSerLyCybGIuProAlaLeuLybArGIYrCybThVaI	725
Db	2134	CAAGGAAATTTGACTGATTAATCTAATGTGAACCTGCATTGGAACAGATTTGTACAGTA	2193
QY	726	AlaGIyAnVaIaAenAnAlaSerIleSeGIyLeuCyAlaVaIaAenThrLybAapAn	745
Db	2194	GCGGGAACCGAAATTAATGCGCTCAATCAGTGGCTTATGCAAGAGCTAACACCAAGGATAC	2253
QY	746	SeGIlybSeSerAapGIuAapAlaArGIyGIuLeuCybGIuLybSeuVaIlybGIuVaI	765
Db	2254	TCTGAAAGAGTGAATGAGAGTGTAGAAAGAACTCTGTGAAGATCAGTAAGAAAGTGTG	2313
QY	766	GIuGIuGIInCybLybAlaLeuProThrGIuLeuGIyGIInProAlaAlaAapLeuLybLyb	785
Db	2314	GAAGAACAAGTCCAAAGCATTTACCAACAGATTTAGACAAACCGCAGCTGATCTTAAAAA	2373
QY	786	AapTyrLybThrTyrGIuGIuLeuLybArGIaGIuGIuAlaAeAenLybSeSer	805
Db	2374	GATTATTAAGACATATGAGAACTTAAAGAAACGTCAGAGAAAGCAAGTAAGCAAGTCCAGT	2433
QY	806	LeuValIleuSeSerLeuIleLybLybAenGIuSeSerAnAlaSerLybSeSerAnSeSerLybAn	825
Db	2434	CTTGTTTTGTCCATCCATTAGAAAAAGAAAGTATATATCAAAAAGTATATAGCAAAAC	2493
QY	826	LybAapLybAenAlaVaISeSerAenGIyLeuGIInAapThrThrLybShISeVaIlybIleu	845
Db	2494	AAGATTAAGATGCCCTTCAAACGCACTTCAGATTCACCAAAACATGTGAAATATCTA	2553
QY	846	ArGIaArGIyAlaLybAapVaISeSerValThrGIuLeuGIuAlaLybAlaAeAapLeuAla	865
Db	2554	CGGAGAGAGTTAAGATGTATCCGTAAACAGATTAAGACTAAACGATTTGATTTGGCA	2613
QY	866	AlaGIuAlaAeGIyArGIyArGIyAlaAapLeuLybGIuArGIyCybAenLybLeuGIuSeSerAap	885
Db	2614	GCAGAAAGTATTTGAAAGATATGTAGATTTGAAGAAAGATGTATTAATTTGAAATCATGAT	2673
QY	886	CybArGIleLybGIuAapCybLybAapLeuGIuGIuAlaCybLybLybIleAenLybAla	905
Db	2674	TGCGAATTTAAGGAGGATTCGAAAGACTTGAAGAGATTTGCAAAAAGATTAATTAAGCT	2733
QY	906	CybArGIaenLeuLybProLeuGIuValLybProIleGIuThrValThrGIuSeSerThrThr	925
Db	2734	TGTGCAATCTGAAGCCCTCTGAGGTGAAGCCGACCAAGACAGACAGAAAGGTAAACG	2793
QY	926	ThrThrThrThrThrThrThrThrThrValAlaAapProLybAlaThrGIuCybLybSeSerLeu	945
Db	2794	ACAACTCACAAACAAACAAACAAACCGTGGCGATCCGAAGCAACAGATGAATCCCTTA	2853
QY	946	GIInThrThrAapThrTTPValThrGIInThrSeSerThrShISeThrSeSerThrIleThr	965
Db	2854	CAGACACAGACACATGGGTTACACAGACATCGACACACACACACAGCAGCTCATCATACA	2913
QY	966	SeThrThrIleThrSeSerLybIleThrLeuThrSeThrThrArGIyCybLybProThrLybCyb	985
Db	2914	TCTACCATCACATCAAAAAATACATTGACATCAACGAGGCGATCAAAACCAACCAAGTGT	2973
QY	986	ThrThrGIyGIuGIuAapAapAlaGIyAapValLybProSeSerGIuGIyLeuArGIyMeSeSer	1005
Db	2974	ACGACAGCGG-----GATGATGACGAAGAGACCTGAAGCAAGTGAAGCTTGAAGGTTAGC	3027
QY	1006	GIyThrAenValMeArGIyValIleValAlaMeValIleSeSerPremeIle	1023
Db	3028	GAGTGGAATGTGAATGAGGGGGGTATGTACCAATGGTATTTGTCATGATTT	3081

	US-10-654-416-14 (1-1023)	US-10-654-416-5 (1-3090)
QY	1 MetAlaArgAlaValAlaArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluGlnHis	20
Db	1 ATGCGCGGGCGGTCAACGCGCGGCAAAAGGTGCACAGAAATGACATTGATGAGAGCAT	60
QY	21 ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLysLeu	40
Db	61 GTTTTGGCTTTGATTTTAAAAAATGAGATTAGAAATACAAATGCAAAACTTAAGTTG	120
QY	41 LysLysLysCysGluGlnGluLeuThrGlnAlaLysLeuAsnIleGluGlnValHisArgLys	60
Db	121 GAAGAAATATTGGCAAAACATTAAACAAATGCAGATTAAATCCAGAAATAAGTTCCAGAAAA	180
QY	61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn	80
Db	181 TTTAAAGATTTCGTGATTAACGGAAACGAATGAAAAATGTCAAGATCTAAATAAAACAA	240
QY	81 IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIle	100
Db	241 GTCAATCAAAAAATGCAATTAATTTCAAGAAAACTTCAACACGCTGCGTGAATAAAATTT	300
QY	101 GlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGlnGly	120
Db	301 TCAGAAATTAAACAGATGAGATTGCAAAAAAGATGAAACAACATGCTATTTTGGAGGGA	360
QY	121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys	140
Db	361 GCATGTCCAACAGAACTTAAAGATGATGCTGCATTAATTAAGGAATAACTGTTATCAAAAA	420
QY	141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly	160
Db	421 GAACGGAACAAATGCGCAAAAGAGTTCTTTGAGGGCGCTTCGCGTGATCTCAATGAA	480
QY	161 SerValIleCysGluLysLysLeuLysGluIleCysPheIleMetGlyArgGluSerAsp	180
Db	481 ACAAGACATGTAAAAAAGCTGAAGAGATTGCGGAATTTAGAAAGAGAAAGCAT	540
QY	181 GluLeuThrAsnLeuLysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLys	200
Db	541 GAATTTAAAGGAGCTTTGCTCTTTATCAAAAAACAATGCTGAAGCTTGTAAACAAAGGA	600
QY	201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaIleLeuGlySerPheLys--	219

DB 601 AAAAGTAATGTGATCTCTTGA AAAAAGAAGTTGAAGAAGCACTTAAAGAAGATGAATTG 660  
 QY 220 LysGIunThrCybLeuGIunLeuGIunGIunGIunCybTyrrPheTyrrIleGIyancCybIyAsp 239  
 DB 661 CGAGAAAATATGCTACTATTACTTGAAGCAATGTACTTCAACAGAGGGAACTGTCAAGCA 720  
 QY 240 AspAsp----- 1leIleIyAsp 245  
 DB 721 GACCAATCAAAAGTCGAATTAACCTTAATATAAAGCTGCMAAAGAAATATATTACCAGGTGT 780  
 QY 246 IleGIunLeuGIyGIyLeuCybGIunGIunGIunIleAlaTyrrMePProGIyPProAsp 265  
 DB 781 GATGATTAATGACGAAAAGTGTGAAAAGAAAATATTGTTTAATATGATCATCAGATCCGAT 840  
 QY 266 PheAspProThrArgProGIunAlaThrIleAlaGIunAspIleGIyLeuGIunGIunPheTyrr 285  
 DB 841 TTCATTCACATTAAGCCAGAGCCTCACTAGCAGAGACATAGGGCTGGAGAAGCTTTAT 900  
 QY 286 LysIyAspValGIunGIunAspGIyAlaPheIleGIyIyAspHisIleuAlaArgAspAlaThrAla 305  
 DB 901 AAGGGGAGAGAGAGATGGAATTTTGTGGAAACAACAATGTAAGATGCACACAGCT 960  
 QY 306 LeuIleuAlaIleuIleGIunAspSerSerIleuIyIyAspAspAspLysGIunIyCys 325  
 DB 961 TTGTTGGACACTACTT-----CTTAAAGAAACCCTTAAAAAGAAAGATGT 100  
 QY 326 GIunGIunAlaLeuGIunIySerCybIyAspAspProHisGIunIleGIunAlaLeuGIunSerLeu 345  
 DB 1006 ATAAAGCCCTTAAAAAAAGCTGCAGAAAACCTCATGAAACATGAGGCTTAGAAAATCTA 1060  
 QY 346 CybIyIyIyAspAnGIyLeuSerAspAspGIyThrIyIyIyCybGIunGIunGIunAsp 365  
 DB 1066 TGTAAAGAAAATTAACCAAGTATGTCATGCAAGAAAATATGATCACTAGCAAAAGAT 1120  
 QY 366 IleAsnIyThrCybIyIlePheThrSerIyAspValThrAsnAspArgIleuPheAspPro 385  
 DB 1126 GTTAACAAACTGTGACAGTCTTACATCAACAAATCTTAAAAACCGTCTTACATTTCA 1180  
 QY 386 ThrIyAspIleAsnAsnGIunIleValGIyrrGIunGIyLeuProThrPheLeuSerAsnGIun 405  
 DB 1186 CCTGATGA-----ATTCCGGAATGGGGAATAATTAACGACATTTCTTATGATGA 1230  
 QY 406 AspCybAlaIyLeuGIunSerTyrrCybPheTyrrPheGIunIyIyCybProAspGIyGIun 425  
 DB 1237 GATGTGCMAAACTGAATCTTATGCTTTTATTAAGAAACTGTTCACAGATGTCAAA 1290  
 QY 426 AsnAlaCybIyAspAsnIleArgAlaThrCybTyrrIyAspArgGIyLeuAspAlaArgAlaAsn 445  
 DB 1297 GAAGCTGTATGAAGTGAAGGACGACCGTGTATAAGAGGGGCTTGATCAGCGGCAAC 1350  
 QY 446 LysValIleuGIunGIunAsnMetArgIyIleuHisGIySerAsnIySerTrpLeuGIun 465  
 DB 1357 AGTGCTGTGCMAAAATATATGCGAGGGTTATTCATGCTGCCTCAATTAAGATTCGTTAAG 1410  
 QY 466 LysPheGIunGIunLeuValIyAspValCybGIunIyLeuIyAspGIunAsnIyGIySer 485  
 DB 1417 AAATTTCACACAGATTTAGCAAAAAGTATGTGAGAAACTS---AAAGGAATATAAGGAAGT 1470  
 QY 486 PheSerAsnAspGIunLeuPheIleIleuCybValGIunProAlaIlyAlaAlaArgLeuLeu 505  
 DB 1474 TTCTCGAAGCATGAATTTGTTGTGTATATCAACACGCAAGGACACACGACTTACTT 1530  
 QY 506 ThrHisAspLeuIyAspMetIyThrIlePheLeuArgGIunIleuAspGIunIyAsArgAsp 525  
 DB 1534 ACACATCAACATCAATATAGATATTTCTTTTACGACAAACACTGGATCAAAAAGAGAT 1590  
 QY 526 PheProThrAspIyAspAsnCybIyGIunGIunGIyArgIyCybGIunAspLeuGIyIyAsp 545  
 DB 1594 TTTCGGAAGATTAAGAGCTGCAGAGAAATTAGGAGAAAATGCCAAGATTTAGCAAGAT 1650  
 QY 546 SerIyAspGIunIleThrTrpProCysHisThrIleuGIunGIunIyAspAspArgLeuGIyThr 565

```
Db 1654 TCAAAAGAAATTACATGGCCATGTCATACACTAGAACGCAATGCAATCGCTTAGGATT 1713
Oy 566 ThrGluIleuLysGlnValIleuLysAspGluIleuLysAspThrLysAspGlnGlu 585
Db 1714 ACAAGAAATTTTAAACAGATTTTATGATGACCAAGAAATATCTTGAAAGATCATGAA 1773
Oy 586 SerCysValIleuLysGlnLysCysAsnLysTTPSerLysArgArgGlyAspAspArg 605
Db 1774 AACGTGCAAAATATTTTAAAGAAAGAAATGCAATTAATGCTGAGAAAGGGGTGATGATGT 1833
Oy 606 PheSerPheValCysValPheGlnAsnAlaThrCysGluLysMetValLysAspValLys 625
Db 1834 TTTTCTTTGATGATGCTTTTCCAAACCTCATCTGATGCTGATGCTGATGCTGATGCTGAA 1893
Oy 626 AspArgCysGluValPheLysLysAsnLysLysAspLysSerTyrIleIleGluPheLysGlu 645
Db 1894 GATAGGTGCAAAATTTGCAAGAAATATGCAAGCATGATGATTAATGATTCCTTTAA 1953
Oy 646 AsnAsnThrAsnLysIleThrThrLysGluLysAsnCysProSerTyrPheIleThrTyrCys 665
Db 1954 AAAAATCAAAATATAAAGCAAGATTCAGCAGCAAAATATTTGCTCCATGCAATCATACAC 2013
Oy 666 AsnArgPheSerProAsnCysProGlyLeuThrLys---GluAsnSerCysThrLysIle 684
Db 2014 GATGATTTTAACTCAATGCTCTGATCTTAAGAAAGCAAAATCTTCTGCAAAATCTT 2073
Oy 685 LysLysHisArgGluProPheTyrLysArgLysAlaLysGluAspAlaLysLysValGlu 704
Db 2074 AAAAATTTTGGCAACATTCCTACAAAGAAAGTTTAAAGATGCTCTTAAGATGAG 2133
Oy 705 LeuGlnGlyLysLeuThrAspLysSerLysCysGluProAlaLysLysArgTyrCysThr 724
Db 2134 CTTGAGAGAAATTTAAGATTAATACTAAATGCTGAACCTGCAATTAAGAAAGATTTGTC 2193
Oy 725 ValAlaGluAsnValAsnAsnAlaSerLysSerGlyLysCysLysAlaAsnThrLysAsp 744
Db 2194 GTATTGAAAGACGTAATTAATGCTGCAATCAGCACTTATGTAAGATTAATCCAAAGT 2253
Oy 745 AsnSerGlyLysSerAsp---GluAspAlaArgLysGluLysCysGluLysLysValLys 763
Db 2254 AAAACTAAAAAGCCGATTAATAATAATGTAAGAAAGCTTGTCTAAATTTAGTGA 2313
Oy 764 GluValGluGluGlnCysLysAlaLysProThrGluLysGlnProAlaLysLysLeu 783
Db 2314 GAGGTGCAACGCAATGCAAGATTAATCAACAGAAATTAACAGAGCTGAAAAAGTCTA 2373
Oy 784 LysLysAspTyrLysThrTyrGluGluLysLysAspArgLysAlaGluAlaMetAsnLys 803
Db 2374 AAAAAGATGTTAAGCATATGAGAACTTAAGAAAGGCAAAAAAGCAATGAAACAG 2433
Oy 804 SerSerLysValLysSerLysLysLysLysLysLysLysLysLysLysLysLysLys 823
Db 2434 TCCAGCTTTGTTTATCACTTGTTAAGAAAAAGAAATTAATCAAGAAATTAATGAGC 2493
Oy 824 LysAsnLysAspLysAsnAlaValSerAsnGlyLysGlnAspThrThrLysHisValLys 843
Db 2494 AAAAACAAGGATTAAGAAATGCTGTTCAACGAGCTTCAAGATTAACCAAAATATGTGAA 2553
Oy 844 IleLysAspArgGluValLysAspValSerValThrGluLysGluAlaLysLysLysAsp 863
Db 2554 ATACTACAGAGAGGCTTAAGAGGCACTTGAACGAATCTGAAGCCAGCAATTTGAT 2613
Oy 864 LeuAlaGluValPheGlyArgTyrValAspLysGluLysArgCysAsnLysLysGlu 883
Db 2614 TTGGCAGAGCAAGGTGTTGAAAGATATGAGCTTGAAGAAATATGAGAAATTTGACT 2673
Oy 884 SerAspCysArgIleLysGluAspCysLysAspLysGluValCysLysLysIleAsn 903
Db 2674 TCGGATTTGGGATTAAGACATTTGCGATGTTTAAAGAAATGCTGTGCAAAAGATTGAG 2733
Oy 904 LysAlaCysArgAsnLysAspProLysGluValLysProHisGluThrValThrLysLys 923
Db 2734 AAGACATGTCACGATCTGAAGCTCTGGAGGTGAAGTGCAGTAATATGTCACAGAAAGC 2793
```

```
Oy 924 ThrThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThrGluCysLys 943
Db 2794 ACAACGACGACCAACAGACAAACAACACCGTTACCATCGAAGGCAACAGAAATGCAAA 2853
Oy 944 SerLeuGlnThrThrAspThrThrValThrGlnThrSerThrHisThrSerThrSerThr 963
Db 2854 TCTTACAGACACAGATTCATGAGGCTTACACAGACATCGACACACACAGACGCTTACC 2913
Oy 964 IleThrSerThrIleThrSerLysIleThrLysThrSerThrAspArgCysLysProThr 983
Db 2914 ATCATCTTACATCATCATCAATCAAAATTAACATTCATCAACAGACGCGATGCAACCAAC 2973
Oy 984 LysCysThrThrGlyGluGluAspAspAlaGlyAspValLysProSerGluGluLysArg 1003
Db 2974 AAGGTACGACAGCGG-----GATGAAGCAAGGACATGAAACCGAGTGAAGATTGAAG 3027
Oy 1004 MetSerGlyTyrPheValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db 3028 ATGAGTGGTGAAGCGTGAATGAAGGGGGGTGATGATGCAATGTTATTTCTTATGATTT 3087

RESULT 5
US-10-654-416-9
; Sequence 9, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3030)
US-10-654-416-9

Alignment Scores:
Pred. No.: 7,24e-299 Length: 3081
Score: 3401.50 Matched: 655
Percent Similarity: 76.17% Conservative: 141
Best Local Similarity: 62.68% Mismatches: 208
Query Match: 62.86% Indels: 41
DB: 19 Gaps: 12

US-10-654-416-14 (1-1023) x US-10-654-416-9 (1-3081)
Oy 1 MetAlaArgAlaValLysArgGluAlaGlyThrGlnAsnSer-----IleAspGlu 18
Db 1 ATGCGCGCGCGGCTCAAGCGGCAAGCTGCAAAACATCAGGGGCTAGTGTATGATGCT 60
Oy 19 GluHisValLeuAlaLeuIleLysGluAspGlyLysSerGluGlnGluCysLysLys 38
Db 61 GAAGAAATCTTTTGGCTTAAATTCAGAGAAATAATATATGATTAATGATGCAAAATA 120
Oy 39 LysLeuLysLysTyrCysGlnGluLysThrGluAlaLysLeuAsnIleGluGlnValHis 58
Db 121 GAATTAGAAATAATTTGTAAGACATTAACGATCCAGATTAATAAACCAAGAAAGTTTCA 180
Oy 59 ArgLysLeuLysGlyPheCysGluAspClyLysAlaAspThrLysCysLysGluLys 78
```





```
OY 779 ProAlaAlaAspLeuLysAspTyrLysThrTyrGluGluLeuLysLysArgAlaGlu 798
Db 2350 GAGGAAAAAGACCTTAAAGATGATTTGAAACATTGTAATAAACTTAAACAGGCGAAG 2409
OY 799 GluAlaMetCysLeuLysSerSerLeuValLeuSerLeuIleLysLysAsnGluSerAsnVal 818
Db 2410 AAAACAAATGAATTAATCCATCTTCTTTATCATCTGTTAAGAAAGATGAATAATATACA 2469
OY 819 SerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGlnAspThr 838
Db 2470 TCGAAAAAATAGACGAAAGACACAGATTAAGATACCGTTTCAACGCGCTTCAAGATAC 2529
OY 839 ThrLysHisValLysIleLeuAspArgGlyValLysAspValSerValThrGluLeuGlu 858
Db 2530 ACAGAACTATGAAAAATCTACGAGAGAGATTAAGATGATCCGTAAACGAATCTGAA 2589
OY 859 AlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTyrValAspLeuLysGluArg 878
Db 2590 GCTAAGGCAATTTGATTTGGTACGACAAAGTATTTGGAAGATATCTGACTTGAAAGAAAGA 2649
OY 879 CysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeuGluGluVal 898
Db 2650 TGTATATAATTGCAATTCAGATTGCAGAGTTAAGAGAGATTGCAAGGATTTTGAAGAGATA 2709
OY 899 CysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLysProHisGlu 918
Db 2710 TGTGAAAGATCAACAGAGATATGTTCAAAATTTAAACCACTGAAGATGAAGCGGACGAA 2769
OY 919 ThrValThrGluSerThrThrThrThrThrThrThrThrThrValAlaAspProLys 938
Db 2770 ACAGTACAGAAAGACCAACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 2829
OY 939 AlaThrGluCysLysSerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHis 958
Db 2830 GCAACAGATGCAAACTTTACAGCAACAGATATCATGATATACACACTTCGACACAT 2889
OY 959 ThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArg 978
Db 2890 ACCACACAGTCTACCATCATCTACATCATCAATCAATCAATCAATCAATCAATCAATCA 2949
OY 979 ArgCysLysProThrLysCysThrThrGlyGluGluAspAspAlaGlyAspValLysPro 998
Db 2950 CGTTCACAAACCAACCAAGTATGACAGAGG-----GATGATGACAGAGAGAGTGAAGCG 3003
OY 999 SerGluGlyLeuAspMetSerGlyTyrAsnValMetArgGlyValIleValAlaMetVal 1018
Db 3004 AGTAGGAGATTGAAGATGATGGTGAACCGTATGAGAGGGGGGTGATGATGATGATGAT 3063
OY 1019 IleSerPheMetIle 1023
Db 3064 ATTCTGTTCAATGATT 3078

RESULT 6
US-10-654-416-1
; Sequence 1, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3042
```

```
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3042)
US-10-654-416-1

Alignment Scores:
Pred. No.: 1,11e-240 Length: 3042
Score: 2762.50 Matches: 541
Percent Similarity: 69.10% Conservative: 177
Best Local Similarity: 52.07% Mismatches: 280
Query Match: 51.05% Indels: 41
DB: 19 Gaps: 18

US-10-654-416-14 (1-1023) x US-10-654-416-1 (1-3042)
OY 1 MetAlaArgAlaValLysArgGluAlaAlaGlyThrGlnAsnSerIleAspGluGluHis 20
Db 1 GTGGCGCGCGCGGTTAAAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 57
OY 21 Val-----LeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluGluCysLysLys 38
Db 58 GTGGGCTTTTGGCTTTAATCTAAAGATTTCTAAAGATGATTAAGATTAAGATTAAGATTAAG 117
OY 39 LysLeuLysLysTyrCysGluGluLeuThrGluAlaLysLeuAsnIleGluGluValHis 58
Db 118 AAATTTAGAAAAAACTTCCAAAGATTAAGTGAACCAATCTAACTCCAGAAACAAAGTACAT 177
OY 59 ArgLysLeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLys 78
Db 178 GAAAGTTTAAAGATTTCTGTGATACGAAAAACGCTGATTAAGATTAAGATTAAGATTAAG 237
OY 79 AlaAsnIleGluLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLys 98
Db 238 AAAATGTTGAAAAAAATGCGGATTTTAAACAGATTAAGAAATGCTGAAAAAG 297
OY 99 LysIleGluIleIleThrAspLysAspCysLysLeuGluGluGluGluGluGluPheLeu 118
Db 298 GAACCTTCAAAATTTGAAAAATGATGAGTGACAAATAATGAACACAGTCTGTTTAA 357
OY 119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLysAsnLysCysTyr 138
Db 358 GAAGACAGATCTCT---GATCTTACAAAGATTTGCAACGATTTTAAGAAACAAATGTTAT 414
OY 139 GlnLysLysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArg----- 155
Db 415 CAGATTAAGCGTGATTAAGTGAAGAAAGTCTTTTAAGATAATTAATTAAGAAAGAAAT 474
OY 156 ---SerAspLeuAsnGlySerValIleCysGluLysLysLeuLysGluIleCysProVal 174
Db 475 TTAAAGATTAATAAATTTCA-----TGTGAAAAATTAACGGAATTAATCTGTCACAGAA 525
OY 175 MetGlyArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGluLysGluThrCysLys 194
Db 526 TTAAGTCAAAATGACGTACGAATTTGATGAATTAATGTTTGTATCAAAAAAATACGTTGAT 585
OY 195 AsnIleLeuIleGluLysAspLysCysGlyThr-----LeuLysThrAspVal 211
Db 586 AATCTTGAAGAAAGAACGACCAAAAGTGTAATCTTTCAAAATCTTAAACGGAATTT 645
OY 212 SerAlaLysLeuGlySerPheLys---LysGluThrCysLeuGluLeuLeuGluGluCys 230
Db 646 AAAACAAATTAAGAGATGAACCACTAAAAAATGCCCCATTAATTAATGAAGAAAGTC 705
OY 231 TyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeuGlyCys 250
Db 706 ATTTTATGATGAAGATGAGTGAAGATTTCA---CTGAAGTGAAGATTTGAAAAA 762
OY 251 LysCysGluGluGluAsnIleAlaTyrMetProGlyProAspPheAspProThrArg 270
Db 763 AAATGTCAAGGAAAAATATTACTTACCATTAATCAATTAATCAAGGTTTGAATCTATACAA 822
```



Db 2926 TGTACGACGAGGAGGAGATGACAGAGAGGTGAAGCCGAGTGAAGGCTGAGATG 2985  
Qy 1005 SerGIYTPAaenValMeArGIYValIleValAlaMeCValIleSerPheMeIle 1023  
Db 2986 AGTGGGTGAGATGATGATGAGAGGGGTGTATTATGCAATGATGATTTTCATTCATGAT 3042

RESULT 7  
US-10-654-416-3  
; Sequence 3, Application US/10654416  
; Publication No. US20040043409a1  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; FILE REFERENCE: glycoprotein (MSG) gene of human Pneumocystis carinii  
; CURRENT FILING DATE: 2003-09-02  
; PRIOR FILING DATE: 2003-09-02  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; PRIOR FILING DATE: 1998-08-17  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 3006  
; TYPE: DNA  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3006)  
US-10-654-416-3

Alignment Scores:  
Pred. No.: 4,356-208 Length: 3006  
Score: 2404.50 Matches: 479  
Percent Similarity: 65.37% Conservative: 195  
Best Local Similarity: 46.46% Mismatches: 320  
Query Match: 44.44% Indels: 37  
DB: 19 Gaps: 19

US-10-654-416-14 (1-1023) x US-10-654-416-3 (1-3006)

Qy 1 MetAlaArGaIaValIleArGIaIaIaIaGIYThrGlnAenSerIleAspGluGluHis 20  
Db 1 GTGGCGCGGCGGCTCAAGCGGCGGCTGCA--GCACAGAAATGATGTTGAAGAAATAT 57  
Qy 21 ValIleuAlaIleuIleuLeuysGIuAspGIYLeuSerGIuGlnGluCyluYlsYlsLeu 40  
Db 58 CTTTGGCTTTCATTTTGAATAATGATGAATAATGATGAATAATGATGAATAATGATGA 117  
Qy 41 LysIleuYrYrCysGIuGluLeuThrGIuAlaIuYlsLeuAenIleGIuGlnValHisArgIys 60  
Db 118 AAAGAGTATTGGAACTTTTAAAAAATGTAACAAAAGAAACAAAACCTAGAAAGAAAG 177  
Qy 61 LeuIysGIYpHeCysGIuAspGIYlsAla--AspThrIysCysLeysGIuLeuIysAla 79  
Db 178 TTAGACGGAATCTGCAAGATGATTAACAATAGAAAGCAAAATGCAAGAAATCAGAAACA 237  
Qy 80 AenIleGIuYlsYrCysThrThrIleYlsGIYlsLeuYlsGIuAlaIleIleYlsYls 99  
Db 228 AAGGTTAAAGCAAAATGTAAGTTTTCAAACAGAACTTGATGAAGCTCAAAAAGGGA 297  
Qy 100 IlegIuIleIleThrAspIysAspCysIysGIuAenGIuGlnGluCysLeuPheLeuGlu 119  
Db 228 GCTTCAACCTTGAAGATATGATTTGAAGAAATGAACGCAATGCTGTTTTCGAG 357  
Qy 120 GIYValIysSerIysGIuLeuIysAspAspCysAenThrIleuArgAenIysCysTyrGln 139  
Db 358 GGAGCATGTCAACAGAACTTAAGATTAATGTAATGAACTGAGGAATTAATGTTATCA 417  
Qy 140 LysIleYlsAlaIysAspIysValAlaGIuGluValIleuLeuArgAlaLeuArgSerAspLeuAen 159

Db 418 AAAAAGCAGACGACGTAGAGAAAAGCTTTTAAAGATGATTTAAGGAACTTTAAG 477  
Qy 160 GYSeValIleCysGIuYlsYlsLeuIleGIuLeuCysProValMeGIYArgGluSer 179  
Db 478 GATAAAAACATGCATAAAATTAAGTTAAAGGGGTTTGTCAAGAAATTCACAAAGAAAGT 537  
Qy 180 AspGIuLeuThrAenIleuCysLeuAenGIuYlsGIuThrCysIysAenIleuIleGIu 199  
Db 538 GATAGCTTAATTAATTAATGCTTGAAGAAAAAAGCGTGGAGATCTTGTATCTAAG 597  
Qy 200 LysAspIysCysGIYThrLeuYlsThrAspValSerAlaIleuGIY-----Ser 217  
Db 598 AAAGATATCAATATCAAACTCTCAAGAGAAATTAATGATCTAGCTTGAAGAAAGAT 657  
Qy 218 PheIysGIuThrCysLeuGIuLeuLeuGIuGlnGluCysTyrPheYrIleGIYAsnCys 237  
Db 658 TTATTAAGAAAAATGTTATTAATTTCTGAAGAAATGTTACTTTATGGGTCAAACTGT 717  
Qy 238 GIYAspAspIleIleIleYlsGIleGIuLeuGIYlsCysGIuGlnGluAenIle 257  
Db 718 --GAAACAGATCGCCAAAGTGAAGATTGCAAGCAAAATGTCAAAAGAAATCTC 774  
Qy 258 AlaTyrMeCProPogIYProAspPheAspProThrArgProGIuAlaThrIleAlaGIu 277  
Db 775 GTTATGACGACACGAGGTTCACTTGTATCTTCAAGAAATTAAGATTAAGTTCAGCAAA 834  
Qy 278 AspIleGIuLeuGIuGluThrPheYrIleYlsValGIuIleAspGIYValPheIleGIYls 297  
Db 835 GAATAGACCTAGAAAATGTTACTAGTAAGACAGTGAAGAAAGAAATTCATATTGGAAG 894  
Qy 298 AenHisLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAenSerIleYls 317  
Db 895 CCATCAATTAAGTAAGTGGCTTTTGTGCTATTATTAAGCAAGATGATGCT----- 948  
Qy 318 LysIysAspAspIysGIuYlsCysGIuGluAlaLeuGlnIysSerCysIysAspProHis 337  
Db 949 CAATAATCTTTTAAGATCAATGTGAAGATGTTTAAAAAATGTCGAAACTTTAA 1008  
Qy 338 GIuHisGIuAlaLeuGIuSerIleuCysIysAsnGIYLeuSerAenAspGIYThrIys 357  
Db 1009 GAGCATATTTATTTAAAGATTTATGATGATTAAGACTATCAGTATCAATCAAAAGAA 1068  
Qy 358 LysCysGIuGluLeuGlnAenAspIleAsnIysThrCysIysIlePheThrSerIysVal 377  
Db 1069 AAAGCCGAGAACTTAATTAAGAGCTTA-----ACAACCCGTATTTAACTGTTCTAA 1122  
Qy 378 ThrAenAsnArgLeuPheAspProThrIysGIYAsnAenGIuIleValGIYTrpGIuGIY 397  
Db 1123 AGGATTGAGAAATTTTTCCTCCAGCT--AATGTAAGGAAATTTATGTTGGCATATG 1179  
Qy 398 LeuProThrPheLeuSerAenGIuAspCysAlaIysIleuGIuSerIYrCysPheYrPhe 417  
Db 1180 TTGCATACATTTCTTGGGAAAGAGAGTGTCAAACTGTTGCGATGTTTATTTATTTG 1239  
Qy 418 GIuYlsYlsCysProAspGIYlsAsnAlaCysIysAenIleArgAlaThrCysTyrIys 437  
Db 1240 AAAAGCCAAAGCTCA--CTTGAAAGCCCTGCATATACTTAAAGCAGCATGTTATAAA 1296  
Qy 438 ArgGIYLeuAspAlaArgAlaAsnIysValIleuGlnIuAenMeCArgGIYMeCLeuHis 457  
Db 1297 AAAGGCTTGAAGCACTGACAAATGAAGCATTTACAAAGTAACTTAAGGAGGAAAAATTGCA 1356  
Qy 458 GYSeAenIysSerIYrPheGIuYlsPheGlnGlnGluValIysValIysGIYls 477  
Db 1357 GCTTCAATATTAACATGCTTGAAGAACTTCAAAAACCTTGTAAGAAAGTTGTGCAAAAG 1416  
Qy 478 LeuIysIysGIuAenIysGIYSerPheSerAenAspGIuLeuPheIleuCysValGln 497  
Db 1417 ACGAAGAGAGAA-----AGTATGATTAATTTTGTACTATGATGAAC 1458  
Qy 498 ProAlaIysAlaIysArgLeuLeuThrHisAspIysArgMeCLeYsThrIlePheLeuArg 517







Qy 373 PheThrSerIyValThrAsnAsnArgLeuPheAspProThrIyGlyAsnAsnGluIle 392  
 Db 2038 -----CAAAGGATATT 2049  
 Qy 393 ValGlyTrpGluGlyLeuProThrPheLeuSerAsnGluAspCys-----AlaLys 409  
 Db 2050 CTAGCCCAAGAGAGCTGAAAAGCTTAT-----AAAGCTGCGATGTCAGAGCCAAA 2100  
 Qy 410 LeuGluSerIyCysPheThrPheGluLysCysProAspGlyGluAsnAlaCysLys 429  
 Db 2101 ACCGAGCT-----GAGAAAAA-----GAATGGAG 2127  
 Qy 430 AsnIleArgIaThrCysTrpLysArgGlyLeuAspAlaArgAlaAsnLysValLeuGln 449  
 Db 2128 AATTTA-----CTCACCCTGAGAGCAAAAAAATTGTA 2163  
 Qy 450 GluAsnMetArgGlyMetLeuHisGlySerAsnLysSerTrpLeuGluLysPheGlnGln 469  
 Db 2164 GAAGAAAGCCAAA-----GAGAGCGTTAAGCGTTATTTGGATTGCGTATCTCAA 2211  
 Qy 470 GluLeuVal-----LysValCysGluLysLeu-----LysLys 480  
 Db 2212 GCCAAAACCGAAGCTGAGAAAAAAGATGCGAATAATTGCTCACCCCTGAGAGCAAAAAA 2271  
 Qy 481 GluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLys 500  
 Db 2272 AAGTTAGAAGAAGCTAAAAAAGCTTAAAGCTTAACTTGAATTCGTTCAAGAGCTAAG 2331  
 Qy 501 -----AlaIleArgLeuLeuThrHisAspLeuArgMetLysThr 513  
 Db 2332 AATGAAAAAGAAAAAAGAAATGCGAATAATTGCTCACCCCTGAA-----GCGAAAAA 2385  
 Qy 514 IlePheLeuArgGlnGlnLeuAspGlnLysArgAspPheProThrAsp-----Lys 530  
 Db 2386 CTTTATAGACAAAGCAAGCTAGATTTGTTGAAAAACCTTAAACCGATTAAGAAAGCAAAA 2445  
 Qy 531 AsnCys-----LysGluLeuGlyArgLysCysGlnAspLeuGlyLysAsp 545  
 Db 2446 AAGCTTTGAAGAAGTCCCTTAAGACTTGCAGAAAAAGCTT-----TTAGCTAAAGAA 2499  
 Qy 546 SerIyGluIleThrTrpProCys-----HisThrLeuGlnGln 559  
 Db 2500 AGCGTTAAAGCTTACTGATGCGTATCTCAAGCCAAAACCTGAAGCTGAGAAAAAAGAA 2559  
 Qy 560 CysAsnArgLeuGlyThrThrGlnIleLeuLysGlnValLeuLeuAspGlnHisLysAsp 579  
 Db 2560 TGCAGAAATTAATCTCACCCCTGAGCGAAGAA-----CTTTTGAAGAAAGCTTAAAAA 2613  
 Qy 580 ThrLeuLysAspGlnGlnIySerCysValLysTrpLeu-----LysGluLys 594  
 Db 2614 AGCGTTAAGCGTTATTTGGATTGCGTATCTCAAGCCAAAACCTGAAGCTGAGAAAAAAGAA 2673  
 Qy 595 CysAsnLysTrp-----SerArgArgGlyAsp 603  
 Db 2674 TGCAGAAATTAATCTCACCCCTGAGCGAAGAAACCTTTAGAGAAAGAAKGCAGAAAGAGC 2733  
 Qy 604 AspArgPheSerPheValCysValPheGln-----AsnAlaThrCys 617  
 Db 2734 GTTAAACCTTACTGATGCGTATCTCAAGCCAAAACGAAAGCTGAGAAAAAAGAAATGC 2793  
 Qy 618 Glu-----LeuMetValLysAspValLysAspArgCysGluValPheLysValAsnIleLys 636  
 Db 2794 GAGAAATTCCTCACCCCTTGAATCGAAAAAAAGTTAGAAAGAAAGTAAAAAGCGTTAAG 2853  
 Qy 637 AlaSerTrpIleIleGluPheLeuGluAsnAsnThrAsnLysIleThrThrLeuGluArg 656  
 Db 2854 GCT-----TATTTGATGCGTATCTCAAGCCAAAACGAAAGCTGAGAA 2898  
 Qy 657 AsnCysProSerTrpHisThrTrpCysAsnArg-----PheSerProAsnCysProGlyLeu 675  
 Db 2899 AAA-----GAATGCGAAAAAATGCTCAAGCGCTGAGAGCAAAAAAAGCTT 2940  
 Qy 676 ThrIyGluAsnAspCysThrLysIleLysLysHisArgGluProPheTrpLysArgLys 695

Db 2941 TTAGAGCAACAACGCGTAGATTGTTGAAAAAGCTTAAACCGAAGCTGATTAAAAAAGC 3000  
 Qy 696 AlaLeuGluAspAlaLeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCys 715  
 Db 3001 TGTGTCAAGAT---CTCCCTTAAGACTTCAGAAAAAGGTTTAAAGCCAAAGAG----- 3051  
 Qy 716 GluProAlaLeuLysArgTrpCysThrValAlaGlyAsnValAsnAsnAlaSerIleSer 735  
 Db 3052 -----AGCTGAAGCTTAT----- 3066  
 Qy 736 GlyLeuCysLysValAsnThrLysAspAsnSerCylLysSerAspGluAspAlaArgLys 755  
 Db 3067 -----AAAGACTGCGTATCAAAAGCTTGAAGTGAAGAAAGAGAAA 3105  
 Qy 756 GluLeuCysGlu-----LysLeuValLysGluValGluGlu 767  
 Db 3106 AAGAAATGCGAAGAAATTAATCTCACCCCTGAGAGCAAAAAACCTTTAGAGAAAGCTTAAAAA 3165  
 Qy 768 GlnCysLysAlaLeuProThrGluLeuGln-ProAlaAlaAspLeu-----LysLysAs 786  
 Db 3166 AGCGTTAAGCTTACTGATTCGCGTATCTCAAGCCAAAACCTGAAGCTGAGAAAAAAGAA 3225  
 Qy 786 PTrpLysThrTrpGluGluLeuLysLysArgAlaGluGluAlaMetAsnLysSerSerLe 806  
 Db 3226 TGCAGAAATTAATCTCACCCCTGAAGCGA-----GAAGACTTGAAGAAAGCTTAAAGAG 3279  
 Qy 806 uValLeuSerLeuIleLys-----LysAsnGluSerAsnVa 818  
 Db 3280 AGCGTTAAGCTTATTAAGAGCTGCGTATCAAAAGCTGAAGATGAAGAAAAAGAAAAAGAA 3339  
 Qy 818 LysLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGly-----LeuG 836  
 Db 3340 TGCAGAAATTAATCTCACCGCTGAGAGCAAAAACTTTAAGCAACAAGCTGATGATGT 3399  
 Qy 836 AspThrThrLysHisValLysIleLeuAspArgGlyValLysAspValSerValThrG 856  
 Db 3400 TTGAAAAACCTTAAGACCGAAGCTGATTAAGAAAAAGGTGTCAAGAGATCTCCCTAAA-GA 3458  
 Qy 856 uLeuGluAlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTrpValAspLeu-- 875  
 Db 3459 CTTCAGAAAAAGGTT-----TTAGCTTAAAGAGAGCGTTAAGCTTATTTGACCTGCT 3512  
 Qy 876 -----LysGluArgCysAsnLysLeu----- 882  
 Db 3513 ATCAAGCTAGCAATGAAGAAAGCAAGAAAGAAAGAAATGCGTACACCCCTGAAGC 3572  
 Qy 883 -----GluSerAspCysArgIleLysGluAspCys----- 892  
 Db 3573 GAAAAAATTTTGAAGAGAGCAAGAGAGCTTAAAGCTTATTAAGAGCTGCGCTGCA 3632  
 Qy 893 -----LysAspLeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysPr 911  
 Db 3633 AGCTAGAAATGAAGAGAA-----AGGAGAGCTTGCAGAAATCTACAC 3677  
 Qy 911 oLeuGluValLysPro-----HisGluThrValThrGluSerThrThrThrThrTh 929  
 Db 3678 GCGTGAAGCGAGAAATCTTTAGAGCAAGAGTAAAGAAAGCACTTAAGCTTATTTGGA 3737  
 Qy 929 rThrThrThrValAlaAspProLys---AlaThrGluCysLysSerLeuGlnThrTh 948  
 Db 3738 CTGCGTATCAAGAGCTAGAAATGAAGAAAAAGAAAGAAATGCGAAGAAATTAATCTACAGCC 3797  
 Qy 948 rAsp 949  
 Db 3798 TGA 3801

RESULT 10  
 US-10-335-977-3219  
 ; Sequence 3219, Application US/10335977  
 ; Publication No. US20040052799A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DOUGLAS SMITH et al



D 2917 -----AAGTTTTCACCAAGAGAGTGTAGGCTTATTGATTCGTACAAAAGCC 2970  
QY AsnAsnGluLeuValGlyTTPGluGlyLeuProThrPheLeuSerAsnGluAspCysAla 408  
D 2971 AAAAACGAA-----GCTGAAGAAAAGAAATGGCAG 3000  
QY LysLeuGluSerTyrCysPheTyrPheGluLysCysProAspGly-----424  
D 3001 AAATTGCTACCC-----CCTGAAGGAGAAAGCTATTTA 3033  
QY 425 GluAsnAlaCysLysAsnIleArgAla-----ThrCysTyrLysArgGly-----439  
D 3034 GAAGAACTTAAGAGAGTGTAAAGCTTAACAAGACTGCGTATCAAGAGCTAGAAATGAA 3093  
QY 440 -----LeuAspAlaArgAlaAsnLysValLeuGlnGly 450  
D 3094 AAAGAGAAACAAAGATCGAGAAATTACTCACCCCTGAAGCTAGAAACTATTAGACAA 3153  
QY 451 AsnMetArgGlyMetLeuHisGlySerAsnLysSerTrpLeu-----464  
D 3154 GAAGTTAAA-----AAGAGCGTTAAGCGTTATTAGACTGCGTTCAAGAGCT 3201  
QY 465 -----GluLysPheGlnGlnGlnLysValLysValCysGluLys---LeuLysLysGlu 481  
D 3202 AGGAATGAAAAAGAGAAACAAAGAA-----TCCGAGAAATTACTACCCCTGAA 3249  
QY 482 AsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLys---500  
D 3250 GCGAGAAACTTTTAGAAGATCAAGCCGTAGAT-----TGTTTGAAAAACGCTAAAGCC 3303  
QY 501 -----AlaAlaArgLeuLeuThrHisAspLeuArgMetLysThrIle 514  
D 3304 GAAGCTGAAAAAAGAGTGTGTCAAGAGATCTCCCTTAAGACTTGCAGAAABAAGTTTAA 3363  
QY 515 -----PheLeuArgGlnGlnLeuAsp-----GlnLysArgAspPheProThr 528  
D 3364 GCTAAAGAGACCGTTAAGCGTTATTAGACTGCGTATCAAGAGCTAGAAATGAAAAAGAG 3423  
QY 529 AspLysAsnCysLysGluLeu-----GlyLysLysCysGlnAspLeuGlnGly 544  
D 3424 AAAAAGAAATGCGAAGTGTGTCTACCGCTGAACGAGAAAACTATTAGAAAGATCTTAA 3483  
QY 545 AspSerLysGluIleThrTrpProCys-----HisThrLeuGlnGln 558  
D 3484 AAAAGCGTTAAGCGTTATTGCGATTCGCTATCAAAAGCCAAAGAAAGCAAGCTGAAAAA 3543  
QY 559 GlnCysAsnArgLeuGlyThrThrGlnIleLeuLysGlnValLeuLeuAspGlnHisLys 578  
D 3544 GAATGCGAGAAATGTCTCACACTGAAGCGAGAAAG-----CTATTAGAAAGAGCTTAA 3597  
QY 579 AspThrLeuLysAspGlnGlnLysSerCysValLysTyrLeu-----LysGlu 593  
D 3598 GAGAGTGTAAAGCTTACAAAGAGCTGCTATCAAGAGCTAGAAATGAAAAAGAGAAACAA 3657  
QY 594 LysCysAsnLysTrpSerArgArgGlyLysAspArgPheSerPheValCysValPheGln 613  
D 3658 GAATGCGAGAA-----CTACTCACCC 3678  
QY 614 AsnAlaThrCysGluLeuMetValLysAspValLysAspArgCysGluValPhe-----631  
D 3679 CCTGAACGAGAGAACTATTAGACCAAGAGTTAAAAAGAGCGTTAAGGCTTATTAGAC 3738  
QY 632 -----LysLysAsnIleLysAlaSerTyrIleIleGlu-----642  
D 3739 TGTGTATCAAGAGCTAGAAATGAAAAAGAGAAACAGATGCGAGAAATTATGACCCCT 3798  
QY 643 -----PheLeuGluAsnAsnThrAsnLysIleThrThrLeuGluArgAsnCys 658  
D 3799 GAACGAGAAATTTTATGAGAAACAGCGCCACAAAGATTAAGCATTAAGATTCG 3858  
QY 659 ProSerTrpHisThrTyrCysAsnArgPheSerProAsn-----671  
D 3859 -----TTGAAAAACGCCGATCTTAAGCAGACAGCGGCTATCATGAG 3900

QY 672 ---CysProGlyLeuThrLysGluAsnSerCysThrLysIleLysLysHisArgGluPro 690  
D 3901 TGTTTGATGTGTTTGAACGATGAAGAGAAAGTCAAAATACCTGCAAGAAAGCAGAGAA---3957  
QY 691 PheTyrLysArgLysValLeuGlnAspAlaLeuLysValGluLeuGlnGlyLysLeuThr 710  
D 3958 -----AAGCGTGTCTTGATTTGTTTGAAGAACGCGCTAGACCGATGAA-----3999  
QY 711 AspLysSerLysCysGluProAlaLeuLysArgTyrCysThrValAlaGlyAsnValAsn 730  
D 4000 GAAAAAAGAAATGCCAA-----AACCTTATAGCGATTTGATCCAAAGAAATCCAA 4050  
QY 731 AsnAlaSerIleSerGlyLeuCysLysAlaAsnThrLysAspAsnSer-----746  
D 4051 AATTTAA-----AAGCAGACAGAACAAATCAATTAAGTAAACA 4092  
QY 747 -----GlyLysSerAspGluAspAlaArgLysGlu 756  
D 4093 GAAAGCTTGCATCAAGCAGACGAGTCTGTGATTAATTAGATGACCTTACTGATCAAGAA 4152  
QY 757 LeuCysGlnLysLeuValLysGluValGlnGlnLysLysAlaLeuProThrGlnLeu 776  
D 4153 GCCATAGAGCAATGTTTAGAGGAGATTGACGATAGCGAAAGGCGCTTAATTCTAGAAATC 4212  
QY 777 GlyGlnProAlaAlaAspLeuLysLysAspTyrLysThrTyrGlnGlnLysLysArg 796  
D 4213 AAACGACAACTGATGAGATGAGATCGGATTTATGCGATCTTAAGAAC---CGCAAAACC 4259  
QY 797 AlaGlnGluAlaIleAsnLysSerSerLysLeuValLeuSerLeuLysLysAsnGlnLys 816  
D 4270 TTTCATTAACATGCGCGCTTAAGGTTATCCATTTGTTTCAATGATTTTAAAAAGTGTGC 4329  
QY 817 AsnValSerLysSerAsnSerLysAsn-----LysAspLysAsnAlaValSerAsnGly 834  
D 4330 GATATTGCTACTATTAAAGCCACTAAATGTTGATGTCGAGCAAAATAGCTAGCATTAATCT 4389  
QY 835 Leu-----GlnAspThrThrLysHisValLysIleLeuArgArgGlyVal 849  
D 4390 ATTATGCTTCCATAGAGCTGATATTACTAGCAATATACGAA---ACAGACAAAAACCAT 4446  
QY 850 LysAspValSerValThrGlnLeuGlnAlaLysAlaPheAspLeuAlaGluValPhe 869  
D 4447 AAGGATTAAGAT-----TTAGAGCTTAA-----TTAGCTAAGGCTTTAGGT 4488  
QY 870 GlyArgTyrValAsp-----LeuLysGlu 877  
D 4489 GCGCATTAAGAAAGATGACGATTAAGAAAAAGTAAAAACCCACAGACAGAAATCTAAAGCA 4548  
QY 878 ArgCysAsnLysLeuGlnLysSerAspCysArgIleLysGlnAspCysLysAspLeuGlnGly 897  
D 4549 GAAGAGAAATAGATAGCAAAAGAT-----GTCCAGAAATCTGCCAAAAATATACAGGAA 4602  
QY 898 ValCysLysLysIleAsnLys-----AlaCysArgAsnLeuLysPro 911  
D 4603 ATGCGCTTAAAGAACAAAAAGAAAGAGATGGGAGTTTGTAGATGAATAATGTAATCC 4662  
QY 912 LeuGlnValLysProHisGlnThrValThrGlnLysThrThrThrThrThrThr 931  
D 4663 ATTCAATATTAAGAAAGAAAGAAAGAAAGAAAGATGAACAGCCCTGTCAACAGCCCTTT 4722  
QY 932 ThrThrValAlaAspProLysAlaThrGlnCysLysSerLeuGlnThrThrAspThrTrp 951  
D 4723 ATAGAGAAAGTATCC-----ACATTT 4746  
QY 952 ValThrGlnThrSerThrHisThrSerThrThrIleThrSerThrIleThrSerLys 971  
D 4747 GTTTTACACAAATGACT-----CCCATTAATAATCACTGACATTTCTTAA 4791  
QY 972 IleThrLeuThrSerThr 977  
D 4792 GTAGATGCCACTGTACAG 4809

```

RESULT 11
US-10-335-977-3220
; Sequence 3220, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 3220:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...5460
; SEQUENCE DESCRIPTION: SEQ ID NO: 3220:
US-10-335-977-3220
US-10-654-416-14 (1-1023) x US-10-335-977-3220 (1-5460)

Alignment Scores:
Pred. No.: 4,086-17 Length: 5460
Score: 311.50 Matches: 247
Percent Similarity: 36.54% Conservative: 179
Best Local Similarity: 21.18% Mismatches: 411
Query Match: 5.76% Indels: 329
DB: 19 Gaps: 61

3 ArgAlaValAlaLeuThrglnaenSerIleAspGluGluHisValLeu 22
1729 AAAGCTTAAAGATCGCATCTCAAGCAGAAAGCTGAAGTGAAGAAAGATGCGAG 1788
23 AlaLeuIleLeuGlu---AspGlyLeuSerGluGlnGluCysGlyLeuLeu 41
1789 AAATTAATCAAGCGCTGAAGCGGAAAAAAGCTTTTGAAGAAAGAGCCAAAGAGCGCTTAAG 1848
42 LysTyrCysGlnGluLeuThrglnaValLeuLeuLeuLeuValHisArgGlyLeu 61
1849 GCTTATTGGATTGCGTATCTCAGGCGCAAAAGCTGAAGCTGAA-----1890

```

```

QY 62 LysGlyPheCysGlu-----AspGlyLeuAlaAspThrLysCysGlyLeuLeu 78
DB 1891 AAAAAAGAAATGCGAAATGCTCAGCCTGACGAGAAAAAGTTAGAAAGAGCCAAA 1950
QY 79 AlaAsnIleGluLysCysThrThrIleGlyLysLeuGluAlaIleLysLys 98
DB 1951 AAGAGCGTTAGAGCTTATTGGATTGCGTATCAAAAGCCAAAAAGAGCTGAAGAGAAA 2010
QY 99 LysIleGln---IleIleThrAspLysAspCysLysGluGlnGlnCysLeu--- 116
DB 2011 GAATGCGAGAAATTAATCAAGCTGAAGCGGAAAAAAGTATTAAGAAATCAAGCGCTAGAT 2070
QY 117 PheLeuGluGlyValCysSerLysGluLeuLysAspCys-----Asn 131
DB 2071 TGTTTGAAAAACGCTAAACCGATGAAGAACGAAAGAGTGCTTGAAGATCTCCCTAAA 2130
QY 132 ThrLeuArgAsnLys-----CysTyrGln 139
DB 2131 GACTTACAGAAAAAGCTTTAGCCAAAGAGCTTAGCGTTATTGGATTGCGTATCA 2190
QY 140 LysLysArgAsp-----LysValAlaGluGluValLeuLeuArgAlaLeuArg 155
DB 2191 AAAGCCAAAAAGAAAGCTGAAGAAAGAAATGCGAGAAATGCTCAGCCCTGAAGCGAGA 2250
QY 156 SerAspLeuAsnGlySerValIleCysGluLysLeuLysGluIleCysProValMet 175
DB 2251 AAGCTATTAGAAAGAGCT-----AAGAAAGCGCTTAAGGCTTACAAAGACTGCGTT 2301
QY 176 GlyArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsn 195
DB 2302 TCAGAGCTAGAGATGAA-----AAGGAAACAAAGATGCGAGAAA 2343
QY 196 Ile-----LeuIleGluLysAspLysLysCysGlyThrLeuLys 208
DB 2344 TTACTCAGCGCTGAAGCGAGAAAGCTATTAGAAAGATCTAAAAAAGCTTAAGGCTTAT 2403
QY 209 ThrAsp---ValSerAlaIleLeuGlySerPheLysGluThrCysLeuGluLeu 227
DB 2404 TTGATTGCGGTATCAAAAGCCAAACGAAAGCTGAAGAAAGATGCGAGAAATGCTC 2463
QY 228 -----GluGln 229
DB 2464 ACCCTGAAGCGAAAGCTATTAGAAAGAGCTAAAGAGCTTAAGGCTTACAAAGAC 2523
QY 230 CysTyrPheTyrIleGlyAsnCysGlyAspAspAspIleLysCysIleGluLeu 249
DB 2524 TGCGTATCAAGAGCTAGAGAT-----GAAAGAGAAACAAAGATGCGAGAAATTA--- 2574
QY 250 GlyLysCysGlnGluGlnaenIleAlaTyrMetProGlyProAspPheAspProThr 269
DB 2575 -----CTC 2577
QY 270 ArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyrLysLeuValGlu 289
DB 2578 ACGCTTGAAGCGAGAAAGCTA-----TTAGAAAGATCTAAAAAAGCTTAAAG 2625
QY 290 Glu-----AspGlyValPheIleGlyLysAsnHisLeu-----ArgAspAlaThrAla 305
DB 2626 GCTTATTGGATTGCGTATCAAAAGCCAAAGAGCTGAAGAAAGAAATGCGAGAAA 2685
QY 306 LeuLeuAla-----LeuLeuIleGlnAsp----- 313
DB 2686 TTGCTACCCCTGAAGCGAGAAAGCTATTAGAAAGAGCTTAAGAGCTTAAGCGCTTAC 2745
QY 314 -----SerSerLeuLysLysLysAspLysGluLysCysGluGluValLeu 329
DB 2746 AAAGATCGCGTATCAAGAGCTGAAGATGAAGAAAGAGAAATGCGAGAAATTAATCTC 2805
QY 330 GlnLysSerCysLysAsnProHisGluIleGluAlaLeuGluGluSerLeuCysLysLysAsn 349
DB 2806 ACGCTGAAGCGAAAGAAAGCTATTAGAGATCAAGCGCTAGAT-----TGTGTAAGAAAC 2859

```







```

Db 2722 AGAAACTCTTGAAGAGCTTAAAGAGCGTTAAAGCTTAAAGC----- 2769
Oy 493 ILeuCyValGlnProAlaLeu-----AlaAlaArgLeuLeu 505
Db 2770 -----TGCCTATCAAAAGCTAAGATGAAAGAAAAAGAAATGCGAGAAATTACTC 2823
Oy 506 ThrHValPheLeuArgMetLeuThrIlePheLeuArgGlnLeu----- 520
Db 2824 ACCGCTGAAGGAAAAA-----CTTTAGAGCAACAAGCTGATGTTGTA 2874
Oy 521 -----AapGlnLeuArg-----AapPheProThrAapLeuAsn 531
Db 2875 AACGCTAAACCGAAGCTGATTAAGAAAAAGTGTCAAAAGATCTCCCTAAAGACTTCAG 2934
Oy 532 CybLeuGlnLeuGlyArgLeu-----CybGlnAapLeuGln 544
Db 2935 AAAAAGGTTTAAAGAGAGCGCTTAAGCTTATTTGACTCGTATCAAGAGCTAAG 2994
Oy 545 AapSerLeuGlnIleThrTrpProCybHleuLeuGlnGlnCybAapArgLeuGly 564
Db 2995 AATGAAAAAGG-----AAAAAGAAATCGCAAAATTGCTC 3030
Oy 565 ThrThrGlnIleLeuLeuValLeuLeuAapGlnHleuLeuAapThrLeuLeuAapGln 584
Db 3031 ACCCTGAAGCGAAAAA-----CTTTAGAAAGACCAAGAGAGCTTAAAGCTTAT 3084
Oy 585 GluSerCybValLeuValLeuLeuGlnLeuCybAapLeuSerAapArgArgGlyAapAsp 604
Db 3085 AAAAAGCTC-----CTCTCTCAAGCTAAGAAATGAAAGAAAGAG----- 3126
Oy 605 ArgPheSerPheValCybValPheGlnAapIleThrCybGln-----LeuMetValLeuAap 623
Db 3127 -----GCTTCCGAAACTTAACTACACCCCTGAA 3153
Oy 624 ValLeuAapArgCybGln-----ValPheLeuLeuAapHleuValAapSerTyIle----- 640
Db 3154 GCGGAAAGAACTTAAAGACAAAGATTAAGAAAGAAAGCTTAAGCTC-----TATTGGAAGCTC 3210
Oy 641 IleGlnLeuLeuGlnLeuAapHleuThrAapHleuThrIleThrLeuGlnLeuAapCybProSer 660
Db 3211 GTATCAAGAGCTAAGAAATGAAAGAAAGAAAGAAATGCGAAATTAATCTCACCCCTGAA 3270
Oy 661 TrpHleuThrTyCybAapArgPheSerProAapCyb----- 672
Db 3271 GCGGAAAGAAATTTTAAAGCAAGAGCTGTAATTTGTTGAAAAAGCTCGAAATGAAAGAA 3330
Oy 673 -----ProGlyLeuThrIleGlnLeuAapSerCybThrLeu 683
Db 3331 GAAAGAAAGCAATGCTTAATAATCTCCCTAAAGACTTAACAGAAATATTTTACCTAAA 3390
Oy 684 -----IleLeuValHleuArgGlnProPheTyIle----- 693
Db 3391 GAGAGCTTAAAGCTTAAAGAGCTGCTCTCAAGAGCTAAGAAATGAAAGAAAGAGAA 3450
Oy 694 -----ArgValAapLeuGlnAapAlaLeuVal 703
Db 3451 GCTTGCAGAAACTTAACTACAGCTGAAAGCGAAAGAACTTAAAGCAAGAAATTAAGAA 3510
Oy 704 GlnLeuGlnGlnLeu-----ThrAapLeuSerLeu 714
Db 3511 AGCGTTAAGGCTTATTTGAGCTGCTATCAAGAGCTAAGAAATGAAAGAAAGAAAGAA 3570
Oy 715 CybGln-----ProAlaLeuLeuArgTyCybThrValAlaGlyAapHleuAsn 730
Db 3571 TGCAGAAATTAATCAAGAGCTGAAAGCGAAATTTTAAAGCAAGAAATCTCAACAAAA 3630
Oy 731 AapAlaSerLeuSerGlyLeuCybLeuValAapHleuThrLeuAapAsn----- 745
Db 3631 GATTAAGCGATCAAAAGAT-----TGCCTGAAAAAGCCGATCTTAAAGCAAGAGCGCTATC 3687
Oy 746 -----SergLeuLeuSerAapGlnAapAlaArgLeuGlnLeu----- 757

```

---

```

Db 3688 ATGAAGTGTGGATGAGTGTGAGCGATGAAGAGAGCTCAAAATCTTCAGAAAGCTAGA 3747
Oy 758 -----CybGlnLeuValLeuValGlnGln-----GlnCyb 769
Db 3748 GAAAGCGCTGTGGATGATGTTGGCTATGCTTAAACCGATGAAGAAAAAGAAATGCG 3807
Oy 770 LeuAlaLeuProThrGlnLeuGlnGlnProAla----- 780
Db 3808 CAAGAACTTATAGCGATTTGATTCAGAAATCCAAATTAAGAAAGACAAACAAACAA 3867
Oy 781 -----AlaAap 782
Db 3868 AATCAATTGATTAACAGAAAGTTGATCAAGCAAGAGAGCTGTTGATTAAGCTTAT 3927
Oy 783 LeuLeuValAapTyIleThrTyGlnGln----- 792
Db 3928 GACCTTACTATCAAGAGCGCTAAGCAATGTTTAAAGGCTTGAAGCGATAGTAAAG 3987
Oy 793 -----LeuValAapArgAlaGlnGlnAla----- 800
Db 3988 GCGCTAATTCAGGAATTAACAGCAAGAGCTGATGATGATTAAGCATCTA 4047
Oy 801 MetAapLeuSerLeu-----ValLeuSerLeuIle 811
Db 4048 AGAAACCGTAAACCTTTGATTAACATGCGCGCTAAGAGCTTATCATTCATTTCAATGAT 4107
Oy 812 LeuLeuAapGlnLeuAapHleuValSerLeuSerAapHleuAsn-----LysAapLeuAsn 829
Db 4108 TTCAAAATAGCGCGCATTTGCCATTAATTAACCCCTAATGTTGATGCGCAAAATA 4167
Oy 830 AlaValSerAapGlnLeu-----GlnAapThrTyHleuValValIle 844
Db 4168 GCTAGCGATTAATCTTATTAATGCTTCATAGAGCTGATATTGCAAGCAATAGAA--- 4224
Oy 845 LeuArgArgGlyValLeuAapValSerValThrGlnLeuGlnAlaValAapHleuAapLeu 864
Db 4225 ACAGAAAAACCAATTAAGAT-----AAGAAATTAAGAGCTAAA-----TTA 4266
Oy 865 AlaAlaGlnValPheGlyArgTyValAap----- 874
Db 4267 GCTAAGCTTATGAGTGCATTAATAAGATGACGATTAAGAAAAAGTAAATAATCCACA 4326
Oy 875 -----LeuLeuGlnLeuArgCybAapHleuGlnLeuAapCybArgIleLeuAapCyb 892
Db 4327 GCAGAACTTAAGCAAGAAACAAATTAAGTACAAAGAT-----GTCGCAAGAACTGCC 4380
Oy 893 LeuAapLeuGlnGlnValCybLeuValAapHleu-----AlaCyb 906
Db 4381 AAGAAATGATCAAGTAAATCGCTTAAAGCAAAAGAAAGAGAGTGGGAATTTGTAGAT 4440
Oy 907 ArgAapLeuLeuProLeuGlnValLeuProHleuThrValThrGlnLeuThrThr 926
Db 4441 GAAATGATTAATCCATTAATGACAAAGAAAGAAAGCAAGAAAGCAAGTGAAGAGCGCT 4500
Oy 927 ThrThrThrThrThrThrValAapProLeuValThrGlnCybLeuGln 946
Db 4501 GTCAAAGAGCGCTTAAAGCAAGAGATGCC----- 4533
Oy 947 ThrThrAapThrTrpValThrGlnThrSerThrThrThrThrThrThrThrThr 966
Db 4534 -----ACATTTGTTTAAAGCGCAATACACC-----CCCATGAATC 4569
Oy 967 ThrIleThrSerLeuIleThrLeuThrSerThr 977
Db 4570 ACTCTGACTTAAAGTAAAGTATGCCACTCTGACA 4602

```

RESULT 13  
 US-10-369-493-25273  
 : Sequence 25273, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052) B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 25273  
LENGTH: 5373  
TYPE: DNA  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-25273  
Alignment Scores:  
Pred. No.: 1,046-13 Length: 5373  
Score: 274.00 Matches: 238  
Percent Similarity: 35.04% Conservative: 193  
Best Local Similarity: 19.35% Mismatches: 453  
Query Match: 5.06% Indels: 346  
DB: 18 Gaps: 51  
US-10-654-416-14 (1-1023) x US-10-369-493-25273 (1-5373)  
6 LysATGtGlnAlaAlaGlyThrGlnAAserIleAspGluGlnHLeuAlaLeuIle 25  
1996 AAAAAAGATTCATATTTCTCAAAAGTTGATGATGAAGACAGATATCTAACTCCGAG 2055  
26 LeuLysGluAspGlyLeuSerLys----- 33  
2056 CTTGATGAAATCGGTTTACCAAAAGTTTATTTACACACTTACTTCAAGTATTATTAAT 2115  
34 -----GlnGluCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 50  
2116 GAAACATATACAGATACGAACCTGATTTCTCAGACCCCTGACGAAAGAGCAATCAAT 2175  
51 LysLeuAAsnIleGluGlnValHisArg-----LysLeuLysGlyPhe 64  
2216 AAAATATCTTTCGAGAAATCGAAAAATTAACAGAGCAATGATACAAAATTTGAAGGT--- 2232  
65 CysGluAspGlyLysAlaAspThrLysCysLysGluLysLysAlaAsnIleGluLysLys 84  
2233 ---GAGATTAATCTTTTCAAAACGAAACCGAAACCCCATGAGAACCTCAACGAAAAA 2289  
85 CysThrTrpIleLysGlyLysLeuLysGluAlaIleLysLysLysIleGlnIleIle--- 103  
2230 TTGATTCGCTTACTTAATGACACAAAGG---TTGATGAGAAATACCAAAATTTGAAT 2346  
104 ---ThrAspLysAspCysLysGluAsnGlnGlnGlnCysLeuPheLeuGluGlnValCys 122  
2347 TCCTCACATTTCTGCTGTAAGAAAAA-----TTTTCCATTTTG 2385  
123 SerLysGluLeuLys-----AspAspCysAsnThrLeuAsnLys 136  
2386 GAAATCAATTCGAAGAACCTGACAGATTCCTTGATGAATGACCAACCTGAGAGATGTA 2445  
137 CysTrpGlnLys---LysArgAspLysValAla----- 146  
2446 CTGGAATTAAGGACAAAGAAATCAACTGCTTTTACTGAGTACAAAGCAATCCAC 2505  
147 ---GlnGluValLeuLeuArgLysSerAspLeuAsnGlySerValIleCysGlu 165  
2506 AAACAAGAAAGACTATCAAACTTTAGAAAAGAGACTTGAACATATTTTGTCTCAAAAG 2565  
166 Lys 185  
2566 AAAAAAGCAAGAGATGCTATTAACAAATGCGCTTAAGATTTATTCCTCTGATGAGAG 2625  
186 CysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLysAsp----- 201

2626 ATGCAAGCACTTGAGAGAAATTTGAAAAATTACAGAGAAAGAAAAAGATTAAGCAATGTC 2685  
202 -----LysLysCysGlyThrLeuLysThrAspValSerAlaLeuGlySerPhe 218  
2686 AACCATCAGAAAGAGACTTAATCACTAAAGAAATTTGCGGCAAAAATTAAGTAATA 2745  
219 Lys-----LysGluThrCys----- 223  
2746 AAACCTATCAATGAAAACTGGAAAGAAATTTCAATGTAAATTAATTTTCAAAAGAA 2805  
223 ----- 223  
2806 AAGGACATATTTGAAAGAACTTTGAGTACAAATCCCGCTTTCAGAGTCATGACAAT 2865  
224 -----LeuGluLeuLeuGluGlnCysTrpPheTrpIleGlyAsnCysGlyAsp----- 239  
2866 CTAGTACGAAACTGAAAGAAAAATTAAGAAATCTTACCAATTAATTAAGATATGCA 2925  
240 -----AspAspIleIleLysCysLysLys----- 247  
2926 GCTGAAATGAGTCTTAAATTAAGCTGTAGAGATCAAAAAAGCAAGCATACAA 2985  
248 -----LeuGlyGlyLysCysGln----- 253  
2986 TTGCTAATTTGCAAAATTAATGATTTATGTCACAGAGAAAAAGAAATTTTCAATA 3045  
254 -----GlnGlnAsnIleAlaTrpMetProGly---ProAspPheAsp 267  
3046 GAAAGAGCAATATGAAAAATATATCAACAAATTAACCAATCTGACTTGAAGAA 3105  
268 ProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTrpLysLys 287  
3106 CAAACGAAAGAGAAATTAATCTCAAAATCCGATCTTCAAAAGATGAATAT----- 3156  
288 ValGluLysAspGlyValPheIleGlyLysAsnHisLeuAspAspAlaThrAlaLeuLeu 307  
3157 ---GAGTCCAGATTAATGCTCTTGAAGAAAAATTAAGAAAGGCGACAAACA----- 3204  
308 AlaLeuLeuIleGluAspSerLeuLysLysAspAspLysGluLysCysGluGlu 327  
3205 -----GCAACGATGAAACGTGAATTAATCTCAGATTTGACTTAAATCAGGGA 3255  
328 AlaLeuGln----- 330  
3256 GAGCTGGAAGCGAGTTAGCAGCTTATTAATAATCTCAAAATGAATCTGAAACAAACTA 3315  
331 -----LysSerCysLysAsnProHisGluHis----- 339  
3316 GAAATTCAGAGAAAGCTTTGAAGAAAGTTAAAGAAACGAGAGCAATTTAAAGAGAG 3375  
340 -----GluAlaLeuGluSerLeuCysLysLysValAsnGlyLeuSerAsn 353  
3376 AAAATTCAACTCGAAGAGAGCCACTGAAACCAACAGACATCAACAGACTTA---CGT 3432  
354 AspGlyThrLysLysCysGluGluLeuGlnAsnAspIleAsnLysThrCysLysIlePhe 373  
3433 GCCAATTTAGATCATTAAGAAAGAGCAGTGAATTTAGCAGCTCACTGAAAGAGATAC 3492  
374 ThrSerLysValThrAsn---AsnArgLeuPheAspPro-----ThrLysGlyValAsnAsn 390  
3493 GAGAGCAAAATTCGCAACAGAAAGCAATATTAACAGAGAAATATTCATTAATATAT 3552  
391 GluIleValGlyTrpGluGlyLeuProThrPheLeuSerAsnGluAspCysAlaLysLeu 410  
3553 GAGATTAATCTCC-----ACTCAACAGAAATTAAGATCA----- 3585  
411 GluSerTrpCysPheTrpPheGluLysLysCysProAspGlyGluAsnAlaCysLysAsn 430  
3586 -----ATTAAGAAAAAATGACGAATCTGAGGGCGAAAGTTAAAGCA 3627  
431 IleArgAlaThr-----CysTrpLysArgGlyLeuAspAla----- 442  
3628 ATGAAGCACTTCAGAGAGAAACAATTTGAATAAGTCAAGAGATTTGATGCTTAAT 3687

```

QY 443 -----ArgAlaenlyValleuGlnleuMetArxGlyMetLeuHleGly 458
Db 3888 TTGCMAATTAAAGCTGAAAAAAGAGAAAGAAAGAAAGCAAGCAAGCAAGTATTGAGTCA 3747
QY 459 SerAnlySerTerPleuGluLysPheGlnGlnleuValLysGlu----- 476
Db 3748 -----ATCAAAAAGTGTAATCAAAAAGGTAAAAATTAAAGATTACAA 3792
QY 477 -----LysleuLysleuGluLysleuGlySer 485
Db 3793 GACCAATGTAATTAAAGAAAGAAAGTGTAGTGAATGGAAACCAATCAGGCTCA 3852
QY 486 PheSerAnlyPheGluPheLeuCyValGlnProAlaValAlaArgLeuLeu 505
Db 3853 GAAGATAGAACTTAATATTG-----GAGCTACAGAGCAATCGAAGATTA 3903
QY 506 ThrHisPheLeuArgMetLysThrLePheLeuArgGlnleuLeuPheGlnLysArgAsp 525
Db 3904 AAGAGAAAGTGAATGCTAAAAAGACTGAATAAAATTCAATTGGAAGAAAGTCAATAT 3963
QY 526 Phe-----ProThrAspLysAsnCyValGluLeuGlyLysCyValAspLeuGly 543
Db 3964 TTATCTAAGGCCAAGAGAAATCAGAACTGATGTTCTAGATTGAAGAGACATCTTCT 4023
QY 544 GluAspSerLysGluLeuThrProCyHisThrLeuGlnGlnCyAsnArgLeu 563
Db 4024 GAGGAAAGAAA-----AATGACAGAGGAGCAATGGAAAAATTG 4062
QY 564 GlyThrGluLeuLysGlnValleuLeuAspGluHisLysAspThrLeuLysAsp 583
Db 4063 AAAAATGAATATACAAATCAAAAATCAGGCTCGAAAAAGAAAGAAAGCTTAATGAA 4122
QY 584 GlnGluSerCyValLysThrLysGlnLysCyAsnLysTerPheArgArgGlyAsp 603
Db 4123 GGATCTTCACAAAT-----ACACAGAAATATCTGAAAAAGATCAATCTTGGAAGAT 4176
QY 604 AspArgPheSerPheValCyValPheGlnAsnAlaThrCyValGluMetValLysAsp 623
Db 4177 GAA-----TTAATTAGGCTACAAATGAAAAC--GAATTAAGAACCAAGAG 4221
QY 624 ValLysAsp--ArgCyGluValPheLysAsnLeuLysAlaSerTyrlleGlu 642
Db 4222 ATGATATAATACAGAGAGAACTAGAGAGCTTGAACCATGAT-----GAG 4272
QY 643 PheLeuGlnAsnAsnThrAsnLysIleThrThrLeuGluArgAsnCyProSerTerPhe 662
Db 4273 CTTTAAAGAAAGAAACAAACACATATAAAAGCTTACAAAGATGAATTTCTTCAATATA 4332
QY 663 ThrThrCyAsnAsnArgPheSerProAsnCyProGluLeuThrLysGluAsnSerCyThr 682
Db 4333 GACAAAATTAATCAAAATGACGAAAAGCTTCAATCAAGACGCAATAT----- 4383
QY 683 LysIleLysLysHisArgGluProPheTyrlsArgLysAlaLeuGluAspAlaLeuLys 702
Db 4384 -----AAACGAGACCTGGAG--AGCTTAAA 4407
QY 703 ValGluLeuGlnGlyLysLeuThrAspLysSerLysCyGlnProAlaLeuLysArgTyrl 722
Db 4408 GAACAACTCCGGCTGACAAAGAAATCCAAAGCTAAAGTTGAAGAAAGGCTGAAAAA--- 4464
QY 723 CyThrValAlaGlyAsnValAsnAsnAlaSerLieserGlyLeuCyValAlaAsnThr 742
Db 4465 -----CTT 4467
QY 743 LysAspAsnSerGlyLysAspGluAspAlaArgLysGluLeuCyGluLysLeuVal 762
Db 4468 GAGAGAAAGATTTGAAA-----GAGAAACAGAGCTGAAAAGTCTAAAGAAATGATG 4521
QY 763 LysGluValGluGluGlnCyValAlaLeuProThrGluLeuGlyGlnProAlaAlaAsp 782
Db 4522 AAAAAAGTAGAGATGAGTGAAGCAATGAACGAGTTGAAATCTCAATGAAAC 4581

```

```

QY 783 LeuLysLysAspTyrlsThrTyrlGluGluLysLysAlaGluGlu----- 799
Db 4582 ATTAAGAAATCAGATGAAAAATTAGAACAGTCCGAAAAAAGTGTGAGGATATTAAA 4641
QY 800 ---AlaMetAnlySerSerLeuValLeuSerLeuLysLysAsnGlnSerAsnVal 818
Db 4642 AACTTGCAATGAGAAAGTCAAGATTGATATCCGAAATCAACGAATCTGAAAAGATATT 4701
QY 819 SerLysSerAnlySerLys-----AsnLysAspLysAlaVal 831
Db 4702 GAAAGACTGAAAGTAAATACAGAGATGAAAGCAAAATCTGCTGTAAGAAACCTTA 4761
QY 832 SerAnlyLysGlnAsnThrThrLysHisValLysIleLeuArgArgGlyValLysAsp 851
Db 4762 AAACAGAGCTTAATACGCAACAAAGAAAATAAGATC-----AACGCA 4806
QY 852 ValSerValThrGluLeuGluAlaLysAlaPheAspLeuAlaGluValPheGlyArg 871
Db 4807 GAAGAAATACAGTTTGAATCTAAATTAGAAAGATATAGACGTGAATCTCAAGACAG 4866
QY 872 TyrValAspLeuLysGluArgCyAsnLysLeuGluSerAsp-----CyArgIle 888
Db 4867 CAGCCGAAATTAAGAT-----AATCAAGAAAGAAAGAAATTAATTAATCTTCCGCTTC 4920
QY 889 LysGluAspCyValAspLeuGluGluValCyValLysIleAsnLysAlaCyAsnArgAsn 908
Db 4921 AAGGAATTAAGACAGAAATTAAGATAGACCCACAGAGGACAAAGAAATCCGAGAAAG 4980
QY 909 LeuLysPro----- 911
Db 4981 CGTAGGCGCCAGTCAAGAAATTTCAATGCAAAAGTCAATGATGAAAAAGCCATG 5040
QY 912 ---LeuGluValLysProHisGluThrValThr-----Glu 922
Db 5041 TTGCTTGAATCAATTAACACAGATTTGTAATTAAGACAGCAAGCAAGTGAAGGATGA 5100
QY 923 SerThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThrGluCys 942
Db 5101 GATACCTTTAAAAAGACACAC-----GATTCACAGACAGAAATT 5142
QY 943 LysSerLeu--GlnThrThrAspThrTrpValThrGlnThrSerThrHisSerThr 961
Db 5143 GAAAGTTGGCTAAAGAACTGATTAACCTGAAGCAAGCAAGCTTAAGTTGAAGAAAGCT 5202
QY 962 SerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArgCyLys 981
Db 5203 AACGAAATCGTTCCGAAATTTGATGACTGATGCTTGGTTACTGACCTAGATGAGAAA 5262
QY 982 ProThrLys-----CyThrThrGluGlu 990
Db 5263 AACGCCAAATATGCTCAAAAGCTGAAGATTTGGGCTTGAAATTAATTCGATGAAAGAG 5322
QY 991 AspAspAlaGlyAspValLysProSerGlu 1000
Db 5323 GATGATGAGAAAGATGATGAAGAAAGAG 5352

RESULT 14
US-10-032-585-6646
: Sequence 6646, Application US/10032585
: Publication No. US20030180953A1
: GENERAL INFORMATION:
: APPLICANT: Terry, Roemer D.
: APPLICANT: Bo, Jiang
: APPLICANT: Charles, Boone
: APPLICANT: Howard, Bussey
: TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
: FILE REFERENCE: 10182-005-999
: CURRENT APPLICATION NUMBER: US/10/032,585
: NUMBER OF SEQ ID NOS: 8000
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6646
: LENGTH: 5641

```





Db 553 TTAGCTAGAAAAATPAGAAAAGAAAGA-----GAAAAGAAAAT 591  
Qy 21 ValLeuAlaLeuIleLeuLys-----GluAspGlyLeuSerGluGln 34  
Db 592 GTTTTAGTAGAGAAATTAAGGATATGAAATATATTAATGAAGATGTTCTTAAGAGAAAGA 651  
Qy 35 -----GluCysLysLysLeuLeuLys 42  
Db 652 AGAGAGCTTTTGAAAGAAAATPAGATTTCTTTAAATGAAGCTTCAAAGAGATCTTAAG 711  
Qy 43 TyrCysGlnGluLeuThrGluAlaLys-----LeuAsnIleGluGln 56  
Db 712 GCAGAGAGGAAATCAATGAAGAAAAGAGCTTTGGCGCTTCAATTTGAATATGAGAA 771  
Qy 57 ValHleArgLysLeuLysGlyPheCysGluAsp----- 67  
Db 772 AAAAATAGATCGTAAGATCTTATGAGAAAAGATGAATAGACTTAAGAGAAA 831  
Qy 68 -----GlyLysAlaAspThrLysCysLysGluLeuLysAlaAsnIleGlu 82  
Db 832 AGAGCCCTTTTAGTGAAGATTCGCGAAGCTTAAGCCTTACATTGATTAATTAGAAAT 891  
Qy 83 -----LysLysCysThrThrIleLysGlyLysLeu----- 92  
Db 892 ACTTAAACAAATTGATATATTAAGAACAAATCCTTCTCGTGAACAACAAATGAA 951  
Qy 93 -----LysGluAlaIleLysLysLysIleGlnIleIleThrAspLysAsp 107  
Db 952 GCTATAAGCTTGAAAAAGAGCATGAGAAAAAATTAAAGTAT----- 996  
Qy 108 CysLysGluAsnGluGlnGlnCysLeu-----PheLeu 118  
Db 997 GCAAGAGTAATPAGAAAAAGACCTCTTAAGTTATGATTAAACATCATTAATATTA 1056  
Qy 119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyr 138  
Db 1057 GATGCAATPAAAGAAAAGATCTTTTGATATATTAAG-----CTTGAAGAAAAGAGCTC 1113  
Qy 139 GlnLysLysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeu 158  
Db 1114 CAAGGAAAAATPAGAAAGCTTAAGTTAGAGCG-----AGT 1149  
Qy 159 AsnGlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGlu 178  
Db 1150 AACCAAGAGATTAATTAACACAGAACATPAAAGATATA----- 1188  
Qy 179 SerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIle 198  
Db 1189 ---GACTCATTTAACTTTGAAAAATACAGAAATTTGGAGAGTAAAGATTGATTAAGCTT 1245  
Qy 199 GlnLysAspLysLysCysGlyThrLeuLysThrAspValSerAlaIleGlnLysSerPhe 218  
Db 1246 CCAGAGAGATTAAG----- 1260  
Qy 219 LysLysGluThrCysLeuGlnLeuLeuGlnCysTyrPheTyrIleGlyAsnCysGly 238  
Db 1261 -----AATPAAATTAATGAAAGAAATATTTTATTAAGCAATTAAT--- 1299  
Qy 239 AspAspAspIleIleLysCysIleGluLeuGlnLysLysCysGlnGlnGlnAsnIleAla 258  
Db 1300 GATGAAAAATTAACATPAAATTAAGCTTGGACTTGATTTGATTAATTTCAAGCT--- 1356  
Qy 259 TyrMetProProGlyPheAspPheAspProThrArgProGluAlaThrIleAlaGluAsp 278  
Db 1357 ---GATTTGAAAAAGCTAAAGTAAAAA----- 1383  
Qy 279 IleGlyLeuGlnGluPheTyrLysValGluGluAspGlyValPheIleGlyLysAsn 298  
Db 1384 -----GAAATGCTTTTAAATTAAGTTAGAGAGAG-----AGAAGT 1419  
Qy 299 HisLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLys 318  
Db 1420 AAATTTAGATCACTACACAAAAAGCTTCAGAGTTTAAATTAAGAGCTTT-----CCT 1470

Qy 319 LysAspAsp-----LysGluLysCysGluGluAlaLeuGlnLysSerCys 333  
Db 1471 AAAGATGATGACTTCTTAACATTTCAAGAAAAATTAATGATTTCTAGCGCAAAAAATGGCGCT 1530  
Qy 334 LysAsnProHisGluHisGluAlaLeuGlnSerLeuCysLys-----LysAsnGlyLeu 351  
Db 1531 AAGTATAGTGAGTATTAATGAAAGCTTAAGGCAAGCTTAAGGCTTGTGTAAGAAATTCACAG 1590  
Qy 352 SerAsnAspGlyThrLysLysCysGluGlnLeuGlnAsnAspIleAsnLysThrCysLys 371  
Db 1591 CAAGTTCTAGAACCAAAAA---GAGCAATATGACAAATTAAGAGATPAAATATCAAAA 1647  
Qy 372 IlePheThrSerLysValThrAsnAsnArgLeuPheAspProThrLysGlyLysAsnGlu 391  
Db 1648 GTT-----AATATTAA 1659  
Qy 392 IleValGlyTyrGluGlnLeuProThrPheLeuSerAsnGluAspCysAlaLysLeuGlu 411  
Db 1660 ATA-----GAGCTTTAGAGAC-----GAA 1680  
Qy 412 SerTyrCysPheTyrPheGluLysLysCysProAspGlyGluAsnAlaCysLysAsnIle 431  
Db 1681 AACATGGCTCATGTACTTAAGAAAGACCTTAAGAGTGAAG---GCATGT----- 1728  
Qy 432 ArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsnLysValLeuGlnGluAsn 451  
Db 1729 ---CCTGATGT----- 1737  
Qy 452 MetArgLysMetLeuHisGlySerAsnLysSerTyrPheGluLysPheGlnGlnLeu 471  
Db 1738 -----GGTTCTGTCATCATTAAGAAAGATTTAAGAAAGTTGAC 1779  
Qy 472 ValLysValCysGluLysLysLeuLysGluAsnLysGly-----SerPhe 486  
Db 1780 CTAAAGGCTTTGAAACACTTAAGAGTGAAGCTCGAAGCTTTGAAAAAAGCTAAGTTT 1839  
Qy 487 SerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAlaAlaArgLeuLeuThr 506  
Db 1840 GAAATATGAGAAATA---GTTATGTGTGAAGCTTAAGCTTAAGTA----- 1881  
Qy 507 HisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAspPhe 526  
Db 1881 ----- 1881  
Qy 527 ProThrAspLysAsnCysLysGluLeuGlnLysArgLysCysGlnAspLeuGluAspSer 546  
Db 1882 ---GAGCAAAAAATTAATTAAGACTTAATGAGAGCATPAAATTAACCTTGGAGAAAGCTTT 1938  
Qy 547 LysGluIleThrTyrProCys-----HisThrLeuGlnGlnCysAsn 561  
Db 1939 AAAGAGATTTCTCTTGAAGATGAGAAAGAAATTTATTAATTTAAAGGAAAAAGTAAAC 1998  
Qy 562 ArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeu 581  
Db 1999 AAATTTAAT-----TTAGAAAAATCAACACTTGATGTATATATTAAGAGACTTA 2046  
Qy 582 LysAspGlnGlnSerCys-----ValLysTyrLeuLysGluLys----- 594  
Db 2047 AGTGAAGCTCAATTAATAATTGAAGTTGAATATCAGAAAGAAAAAGACAGTTGAAGACAG 2106  
Qy 595 CysAsnLysTyrSerAlaArgGlyAsp-----AspArg 605  
Db 2107 TGTGAAAA---CGTATGTTGATTTAAAAAGCAATTGGAAGAGCTATTAAAGAA 2160  
Qy 606 PheSerPheValCys-----ValPheGlnAspAlaThrCysGluLeuMetValLysAspVal 624  
Db 2161 TTATATGAGATGACATTAACATAGAAATTTAAAGCTGAATGAAGATTTCAAGATTTT 2220  
Qy 625 Lys---AspArgCysGluValPheLysLys-----AsnIleLysAlaSerTyrIle 640  
Db 2221 AAATTTGAAATGAAAGAAATATTTGGAAGAAAGAAAGATTAAGGCTTAAGAGCTGA 2280



```
Qy 641 ILeGluPheLeuGluuAsnAnthrAsnLysIleThrThrLeuGluuArgAsnCysProSer 660
    |||::|||
Db 2281 ATAAAGATCTTAAGAAACCTTTAAAC-----ATA 2310

Qy 661 TrpHisThrTyrcybaAsnArgPheSerProAsnCysProGlyLeuThrLysGluuAsnSer 680
    |||::|||
Db 2311 AGACACACAGAAAAAGACAGCTAAATGATTAAGTGTAGTAACTAAAGAGAGACTTACG 2370

Qy 681 CysThrLysIleLys--LysHisArgGluProPheTyrlsAsnArgLysAlaLeuGluuAsp 699
    |||::|||
Db 2371 AAAAATTAAGCTGCTTAAGAAAAAGATTAATAATAACGAAATAATGAACATAATA 2430

Qy 700 AlaLeuLysValGluLeuGlnGlyLysLeuThrAsp--LysSerLysCysGluProAla 718
    |||::|||
Db 2431 AAAAATTAAGTAAAGATTAAGATTAATTAATTAAGACTTAAGAAAAAGATTAAGCGTACA 2490

Qy 719 LeuLysArgTyrcythrValAlaGlyAsnValAsnAsnLaserIleSerGlyLeuCys 738
    |||::|||
Db 2491 ATTAATAATAATAAT----- 2502

Qy 739 LysAlaAsnThrLysAsnAsnSerGlyLysSerAspGluAsnAlaArgLysGluLeuCys 758
    |||::|||
Db 2503 -----GAAAGAACGTAAATTATGT 2523

Qy 759 GluLysLeuValLysGluValGluGluGlnCysLysAlaLeuProThrGluLeuGln 778
    |||::|||
Db 2524 GATTAAGAAAAATGAATGAATTAAGACACAAATACGAAAGTGTAGTGAATAATAAG 2583

Qy 779 ProAlaAlaAspLeuLys-----LysAspTyrlsThrTyrgluGluLeuLysArg 796
    |||::|||
Db 2584 TATCACAGTAAATTGTCCAGCTTCACAGATACGAAAGTT--AATGATATACACAAATTA 2640

Qy 797 AlaGluGluAlaMetAsnLysSerSerLeuValLeuSerLeuLysLysAsnGluSer 816
    |||::|||
Db 2641 AATAAAATACTTAAGAAAGAAATTTGAAATATAGAGAAAGCTAAGAAATTAATCTA 2700

Qy 817 AsnValSerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGln 836
    |||::|||
Db 2701 AACGATTAAGAAATTAATTTGCTTAAGAAAGCATGTGAAAGATTAATAATGAACCTTCA 2760

Qy 837 AspThrThrLysHisValLysIleLeuArgArgGlyValLysAspValSerValThrGlu 856
    |||::|||
Db 2761 AAGTAATATGAGCTGTGAGCTTTTAAGCAAGACTTAAGAAATTAAGAAATTAACGAA 2820

Qy 857 LeuGluAlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTyrlsAlaAspLeuLys 876
    |||::|||
Db 2821 -----GAAAAGTGATGAAATTCAAAAATAATAGAGTTGAGAAAGCT 2862

Qy 877 GluArgCysAsnLysLeu--GluSerAspCysArgIleLysGluAspCysLysAspLeu 895
    |||::|||
Db 2863 TCCAAAGCAAAAGCGCTTCAAGAAAGAAAGTAAACTTGAAGAAAGCTTAATAAACATA 2922

Qy 896 GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys 915
    |||::|||
Db 2923 GAATATAAACTTAAGAACTTGGAACCTTCTTAACCTTAAGCAAGATTAAGAACATATA 2982
```

Search completed: October 29, 2005, 15:51:05  
Job time : 2041 secs

*This Page Blank (uspto)*